# UNIVERSITY OF SOUTHAMPTON

FACULTY OF ARTS AND HUMANITIES

School of Humanities

**Evaluating Machine Translation Post-editing Training in Undergraduate Translation Programs- An Exploratory Study in Saudi Arabia**

By

**Halah MohammadZain Samman**

BA, English-Arabic Translation (Hons); MA, Computer-assisted Language Learning

ORCID ID 0000-0003-2026-2932

Thesis for the degree of Doctor of Philosophy in Modern Languages

January 2022

**UNIVERSITY OF SOUTHAMPTON**

# ABSTRACT

Faculty of Arts and Humanities

School of Humanities

Thesis for the degree of Doctor of Philosophy

**Evaluating Machine Translation Post-editing Training in Undergraduate Translation Programs- An Exploratory Study in Saudi Arabia**

Halah MohammadZain Samman

---

The modern translation industry is using machine translation post-editing (MTPE) widely, and the translation industry in the Arab World is following the global lead. However, while MTPE training is offered in many language pairs around the world, MTPE training in English-Arabic is still not officially offered in translation training programmes in the Arab World. The aim of this study is to evaluate the effectiveness of MTPE training in a female undergraduate translation programme in Saudi Arabia by examining students' opinions about MTPE and comparing its productivity and quality with an established practice in the translation classroom, i.e., human translation (HT). To achieve its aim, this study used a mixed-method design of the 'Kirkpatrick Model of Learning Evaluation'. Focus group discussions and retrospective pre-test surveys were used to examine students' opinions as well as a pre-post experiment which involved two groups of students (29 in the control group and 31 in the experimental group) that was used to compare the productivity of students and the quality of translated texts when using MTPE as compared with HT. Students' opinions that were revealed through the pre-intervention focus group discussions were generally mixed with a preference shown in favour of HT, except for translation speed as most of the students thought that MTPE was the faster method of translation. As for the survey, students' pre-intervention responses supported those opinions revealed in the focus group discussions. However, post-intervention responses revealed a statistically significant shift towards more acceptance of MTPE training and use, indicating that the more students learned about the features of MT and MTPE skills and practiced them, the more positive their opinions became. Statistical results from comparing students' productivity showed a medium effect size which indicates that MTPE cannot be ignored as a method to increase productivity in translation. The effectiveness of MTPE in translation quality was evaluated by measuring error count and error type. Error count analysis indicated that students who used MTPE have increased scores in a similar manner to those who used HT but not more. The analysis of error type showed that while MTPE helped students avoid deletion and technical errors, the number of errors relating to accuracy, comprehension and grammar were more frequent in Arabic MTPE translated texts.

# Table of Contents

**Table of Contents** ..... **i**

**Table of Tables** ..... **vii**

**Table of Figures** ..... **ix**

**Research Thesis: Declaration of Authorship** ..... **xi**

**Thesis Dedication** ..... **xiii**

**Acknowledgements** ..... **xv**

**Definitions and Abbreviations** ..... **xvii**

**Chapter 1 Introduction** ..... **19**

1.1 Rationale for the Study ..... 20

1.2 Personal Motivation ..... 22

1.3 The Aim of the Study ..... 24

1.4 Contextual background of the study ..... 26

1.5 The Socio-technical context for MTPE ..... 27

1.6 Organization of Thesis ..... 28

1.7 Chapter Summary ..... 29

**Chapter 2 Literature Review** ..... **31**

2.1 Introduction ..... 31

2.2 Background ..... 31

2.2.1 The Current Status of Translation Training ..... 31

2.2.2 The Location of the Current Study within Translation Studies ..... 34

2.2.3 Arabic MT: From Statistical MT to Neural MT ..... 36

2.3 Conceptual Framework ..... 37

2.3.1 Key Concepts ..... 38

2.3.2 Effectiveness ..... 38

Opinion 39

Translation Productivity ..... 40

Translation Quality ..... 41

Learning Evaluation Models ..... 42

2.4 Empirical Research on Post-editing ..... 46

2.4.1 Post-editing Definition and Types ..... 46
2.4.2 Current Status of Post-editing Training ..... 47
2.4.3 Opinions ..... 51
2.4.4 Translation Productivity ..... 57
2.4.5 Translation Quality ..... 63
TQA Approaches ..... 64
Studies Comparing Translation Quality Between HT and MTPE ..... 65
Common Errors in Arabic Translation and MT ..... 69
2.5 Summary and Concluding Remarks ..... 71

**Chapter 3 Methodology ..... 73**

3.1 Introduction ..... 73
3.2 Research Summary ..... 73
3.3 Philosophical Assumptions ..... 75
3.4 Research Design ..... 77
3.4.1 Experimental Design ..... 77
3.4.2 Mixed-method Design ..... 80
3.5 Research Setting ..... 84
3.5.1 Designing the Intervention Course and Translation Tasks ..... 85
3.5.1.1 Text Selection ..... 85
3.5.1.2 Test Material ..... 86
3.5.1.3 Metrics for Text Difficulty, Length, and Readability ..... 87
3.6 Research Sample: Translation Students ..... 88
3.6.1 Detailed Selection Criteria ..... 89
3.6.1.1 Students' Knowledge of MT ..... 90
3.6.1.2 Students' Profiles ..... 90
3.7 The Teaching Intervention ..... 91
3.8 Translation Quality Assessment ..... 93
3.8.1.1 Error Type ..... 97
3.8.1.2 Evaluators ..... 97
3.8.1.3 Inter-rater Reliability ..... 98

3.8.1.4 Error Count ..... 99

3.9 Data Collection ..... 100

3.9.1 Focus Group Discussions ..... 100

3.9.2 The Process of Data Collection ..... 101

3.9.3 Translation Productivity ..... 102

3.9.4 Translation Quality ..... 103

3.9.5 Retrospective Pre-test Survey ..... 104

3.9.6 Piloting the Study ..... 105

3.10 Data Analysis ..... 106

3.10.1 Qualitative Data Analysis ..... 109

3.10.1.1 The Process of Thematic Analysis Using Microsoft Excel ..... 109

3.10.2 Quantitative Data Analysis ..... 110

3.10.2.1 Retrospective Pre-test Surveys ..... 111

3.10.2.2 Translation Productivity ..... 111

3.10.2.3 Translation Quality ..... 111

3.10.2.4 Data Cleansing ..... 112

3.11 Ethical Considerations ..... 114

3.12 Reflection on Researcher Positionality ..... 115

3.13 Summary ..... 116

**Chapter 4 Results ..... 117**

4.1 Introduction ..... 117

4.2 Opinion Results ..... 118

4.2.1 Findings from Focus Group Discussions (Pre) ..... 118

4.2.1.1 MT Use: How often do you use MT? ..... 118

4.2.1.2 Degree of Reward: How rewarding is PE compared to HT? ..... 120

4.2.1.3 Usefulness: How useful is MT output according to translators? ..... 123

4.2.1.4 Speed: Which translation method is perceived as being faster? ..... 125

4.2.1.5 Quality: How is the final quality of both methods of translation perceived? ..... 127

4.2.1.6 Preference: Which translation method is the most preferred translation method? ..... 128
4.2.1.7 Summary of Findings from FGDs ..... 130
4.2.2 Results from Retrospective Pre-test Survey (Post) ..... 131
4.2.2.1 MT Use ..... 132
4.2.2.2 Degree of Reward ..... 133
4.2.2.3 Usefulness ..... 134
4.2.2.4 Ease of Use ..... 135
4.2.2.5 Speed ..... 136
4.2.2.6 Quality ..... 137
4.2.2.7 Preference ..... 138
4.3 Translation Productivity Results ..... 139
4.3.1 Productivity Gain in Ga ..... 140
4.3.2 Productivity Gain in Gb ..... 140
4.3.3 Inter-group Productivity Gain (Comparing Ta and Tb of Ga vs. Gb) ..... 141
4.4 Translation Quality Results ..... 144
4.4.1 Error Type ..... 144
4.4.2 Error Count ..... 148
4.4.2.1 Horizontal Analysis ..... 148
4.4.2.2 Longitudinal Gb: Intra-group MTPE Effectiveness ..... 155
**Chapter 5 Discussion ..... 161**
5.1 Introduction ..... 161
5.2 Students' Opinions about MTPE ..... 161
5.2.1 Discussion of Findings from FGDs and Retrospective Pre-test Survey ..... 162
5.2.1.1 Degree of Reward ..... 162
5.2.1.2 Usefulness ..... 164
5.2.1.3 Ease of Use ..... 166
5.2.1.4 Speed ..... 166
5.2.1.5 Translation Quality ..... 167
5.2.1.6 Preference ..... 168

5.3 Discussion of Translation Productivity ..... 169
5.4 Discussion of Translation Quality ..... 171
5.4.1 Error Type ..... 172
5.4.2 Error Count ..... 174
5.5 Chapter Summary ..... 177

**Chapter 6 Conclusions ..... 179**

6.1 Introduction ..... 179
6.2 Thesis Overview ..... 179
6.3 Addressing the Research Questions ..... 182
6.4 Contributions of the Study ..... 184
6.4.1 Guidelines for Arabic MTPE Stakeholders ..... 184
6.4.2 Empirical, Methodological, and Theoretical Contributions ..... 186
6.5 Limitations and Difficulties Experienced During Research ..... 187
6.6 Recommendations for Future Research ..... 189
6.6.1 Allocated Time for the Study ..... 189
6.6.2 Participants and Evaluators ..... 189
6.6.3 MT Systems ..... 190
6.6.4 Methodology ..... 190

**Appendix A Ethical Approval- Southampton University ..... 191**
**Appendix B Focus Group Discussion Questions ..... 193**
**Appendix C Consent Form- English ..... 195**
**Appendix D Consent Form-Arabic ..... 197**
**Appendix E -Participant Information Sheet (PIS)-English ..... 199**
**Appendix F - Participant Information Sheet (PIS)-Arabic ..... 203**
**Appendix G -External Approval ..... 207**
**Appendix H - Visited Research Designs ..... 209**
**Appendix I – Experiment Texts (Ta and Tb) ..... 211**
**List of References ..... 213**

# Table of Tables

Table 2-1 User acceptance: perceived usefulness, perceived ease of use, and their conceptualisation (Davis 1989) ..... 40

Table 3-1 Research summary ..... 74

Table 3-2 Comparability of texts ..... 88

Table 3-3 Topics covered in MTPE training ..... 92

Table 3-4 Processes of collecting data ..... 100

Table 3-5 Summary of data collection- Opinions ..... 101

Table 3-6 Research questions' justifications and units of analysis ..... 107

Table 4-1 List of abbreviations ..... 117

Table 4-2 Identified themes of MT use ..... 119

Table 4-3 Identified themes in the degree of reward ..... 120

Table 4-4 Identified themes regarding usefulness ..... 123

Table 4-5 Identified themes regarding speed ..... 125

Table 4-6 Identified themes regarding quality ..... 127

Table 4-7 Identified themes regarding preference ..... 129

Table 4-8 Summary of students' results ..... 132

Table 4-9 Productivity gain in Ga (Ta and Tb) ..... 140

Table 4-10 Productivity gain in Ga (Ta and Tb) ..... 141

Table 4-11 Inter-group productivity gain ..... 142

Table 4-12 The aims of horizontal analysis ..... 148

Table 4-13 Ta (fs) of Ga and Gb ..... 149

Table 4-14 Tb (fs) of Ga and Gb ..... 153

Table 4-15 Inter-group MTPE effectiveness ..... 154

Table of Tables

Table 4-16 Research questions and methods ..... 156

Table 4-17 Final scores of GbTa and GbTb ..... 159

Table 4-18 Translation quality scores (Ta vs. Tb) ..... 160

Table 5-1 Score differences in all Aspects of Performance for the best scoring students in Gb176

# Table of Figures

Figure 2-1 Holmes's conception of Translation Studies (Munday 2016: 17) ..... 35

Figure 2-2 Munday's conception of 'Applied Translation Studies' (2016: 20) ..... 35

Figure 2-3 Kirkpatrick's Model of Learning Evaluation ..... 45

Figure 2-4 Conceptual framework for the evaluation of the effectiveness of MTPE training ..... 46

Figure 3-1 Kirkpatrick's Model of Learning Evaluation ..... 82

Figure 3-2 Sequential mixed methods (Teddlie and Tashakkori 2006) ..... 83

Figure 3-3 The language lab where data were collected ..... 85

Figure 3-4 Examiners' Mark Sheet ..... 94

Figure 3-5 Calculating the group mean- Example ..... 112

Figure 3-6 Normal distribution curve of Ta ..... 114

Figure 3-7 Normal distribution curve of Tb ..... 114

Figure 4-1 MT use ..... 133

Figure 4-2 Degree of reward ..... 134

Figure 4-3 Usefulness ..... 135

Figure 4-4 Ease of use ..... 136

Figure 4-5 Speed ..... 137

Figure 4-6 Quality ..... 138

Figure 4-7 Preferred method of translation ..... 139

Figure 4-8 Total translation time (Ta vs. Tb) for Ga and Gb ..... 140

Figure 4-9 Inter-group productivity gain ..... 143

Figure 4-10 Values of effect sizes and their interpretation (Cohen 1992) ..... 143

Figure 4-11 Types and number of errors found in GbTa and GbTb ..... 145

Figure 4-12 Error Type and Count in Raw MT Output and MTPE ..... 147

Table of Figures

Figure 4-13 Ta(fs) of both Ga and Gb ..... 151

Figure 4-14 (fs) of Ga and Gb in both Ta and Tb..... 154

Figure 4-15 (fs) of GbTa and GbTb..... 156

# Research Thesis: Declaration of Authorship

| | |
|---|---|
| Print name: | **Halah MohammadZain Samman** |

| | |
|---|---|
| Title of thesis: | **Evaluating Machine Translation Post-Editing Training in Undergraduate Translation Programs- An Exploratory in Saudi Arabia** |

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission;

| | | | |
|---|---|---|---|
| Signature: | | Date: | 27/1/2022 |

# Thesis Dedication

To Mum (1940-2020) and my brother Ibrahim (1966-2021),

We did it!

# Acknowledgements

All the praises and thanks be to Allah, my Lord.

I would like to thank many people who have supported me through the journey of my doctoral research.

First of all, there are no words to express my gratitude to my main supervisor, Dr Ian McCall for his support, encouragement, commitment and for his valuable guidance throughout my PhD journey. This PhD would not have been accomplished without you, Ian!

I wish to thank my second supervisor, Prof. James Minney for his support and cooperation, which I really appreciate. I would also like to express appreciation to my colleagues in the College of Languages and Translation, King Saud University, for all their assistance since the beginning of my PhD.

I would like to thank my fellow research colleagues and all my friends with whom I shared friendship, ideas and much debate during my time at the University of Southampton.

Thanks to my beloved sisters Jehan and Manal for believing in me. Thanks to my dearest friend Noha for her persistent support.

My deepest gratitude goes to my eldest son and my best friend forever, Homiedan. 'Thank you for all the letters of support that you wrote me, all the cups of coffee you offered me, and for all the patience that you taught me'. Thanks to my beloved twins Sulyman and Ibraheem who filled my life with all the love, playing, hugs, and kisses that I needed to persist.

My most heartfelt love and appreciation go to my *Rock*. 'Thank you for believing in me from the get-go, and for all the academic and spiritual support that you showered me with during my toughest times'.

# Definitions and Abbreviations

| | |
|---|---|
| **CAT** | Computer-assisted Translation |
| **COLT** | he College of Languages and Translation |
| **GT** | Google Translate |
| **MT** | Machine Translation |
| **HT** | Human Translation or translating from scratch |
| **ST** | Source Text |
| **MTPE** | Machine Translation Post-editing |
| **PE** | Post-editing |
| **TQA** | Translation Quality Assessment |
| **Ga** | Group A, the control group, translation students in level 10 |
| **Gb** | Group B, the experimental group, translation students in level 10 |
| **Ta** | Baseline pre-test, the pre-test taken by students in both Ga and Gb |
| **Tb** | Study post-test, the post-test taken by students in both Ga and Gb |
| **TT** | Translated Text |
| **TAUS** | Translation Automation User Society |
| **fs** | Final score: the overall score assigned for the test; maximum score is 100 |
| **N** | The number assigned to students as their ID |
| **AoP1** | Aspect of Performance 1: Comprehension, Accuracy and Register |
| **AoP2** | Aspect of Performance 2: Grammar, coherence, cohesion, and organization of work |
| **AoP3** | Aspect of Performance 3: Technical aspects |

# Chapter 1 Introduction

The translation services market has become increasingly decentralised (Int 2005). It is no longer determined by the geopolitical borders of countries. This development, in addition to the growing demand for English to Arabic translations that are characterised as fast and of acceptable quality to the client, have motivated many researchers to study the gap between translation training and the translation market in countries where English-Arabic is the dominant direction of translation. Even though those studies in addition to others that have focused on exploring Arabic machine translation errors claim that technology-related courses would offer a means to close the gap between training and the market, translation training in the Arab World still focuses on improving linguistic competence and generic translation skills, and students tend to consider the university programme a venue for learning a second language to get a job that requires mastery of the English language (Al-Jarf 2017). This might be one of the reasons why translation training in the Arab world has been criticised for not preparing translators for employment in the translation profession (Al-Jarf 2017). Surveys on the contents of academic programmes in the Arab World have highlighted the lack of courses to improve instrumental competence in future translators:

> 'The number of courses offered at both the undergraduate and graduate levels related to developing professional and instrumental competence is relatively small (10% and 22% respectively, of the total number of [surveyed] courses). Moreover, those courses are not offered by all translator training programmes' (Al-Batineh and Bilali 2017: 198)

Machine translation post-editing (MTPE), or the 'term used for the correction of machine translation output by human linguists/editors' (Senez 1998: 289), was suggested as one of the ways to both exploit the technology in favour of translators as well as a method to fix the errors that MT output generates (Al-khresheh and Almaaytah 2018). However, no academic programme in the Arab World that I know of has put these recommendations into practice. Therefore, rather than focusing on exploring more details about the gap between the translation training and the job market or recommending generic improvements to translation training programs, this study intends to specifically evaluate MTPE as a training intervention in an undergraduate translation program.

Throughout this study, I intend to operationalize and evaluate MTPE in an undergraduate translation classroom, so that readers can decide on the effectiveness of MTPE training as one of the means of fixing the recurring errors in Arabic MT and bridging the gap between translation

training and the job market. Then to provide implications for future teaching and research. The study will adopt the Kirkpatrick Model of Learning Evaluation to conduct a thorough evaluation of MTPE (see **Error! Reference source not found.**). In addition to this introductory section, chapter one is comprised of five sections: the rationale for the study, the personal motivation, the aim of the study, the contextual background of the study, and finally, the organization of this thesis, which explains the structure of this study.

## 1.1 Rationale for the Study

The rationale for this study stems from three areas in the literature, which are the previously mentioned gap between translation training and the job market in the Arab World, errors in Arabic MT systems, and the evaluation of MTPE. Firstly, the gap between translation training and the job market in the Arab-speaking world has been highlighted numerous times (e.g. Alotaibi 2014; Alenezi 2016; Abu-ghararah 2017; Fatani 2007; Fatani 2009), and calls for improvements in university-level translation training have been heightened (e.g. Abu-ghararah 2017), yet research on ways to bridge that gap is less emphasised and researchers are still wondering whether the computer-assisted translation (CAT) courses offered in translation training programmes meet the needs of both the future translators and the translation industry (e.g., Al-Jarf 2017).

Secondly, research on Arabic MT systems has repeatedly reached similar conclusions; Arabic MT systems generate output that requires fixing, and the need for post-editing is growing more even with the emergence of neural machine translation (NMT) systems in 2015. A recent study (Al-khresheh and Almaaytah 2018) has confirmed the need for human intervention to fix the Arabic MT output, yet no real hands-on evaluations of the skills that would provide the required level of human intervention on MT output to fix its apparent errors have been conducted.

Finally, studies conducted to evaluate the human skill of fixing MT errors and gaining productivity (i.e., post-editing or PE) have been mainly focusing on language pairs that include (to name a few) English, French, Spanish, Dutch, Catalan, Italian, Turkish, and Chinese. The language pair English-Arabic has rarely been touched upon. Throughout my review of literature, the evaluation of English-Arabic has appeared in two studies only. It appeared in a study that focused on comparing English-Arabic MTPE texts by professional translators with other language pairs (Green et al. 2013), and in a study which evaluated the temporal and technical efforts exerted in MTPE processes performed by non-native speakers of Arabic (Haji Sismat 2016).

Therefore, formative evaluation of the effectiveness of integrating a CAT training course in the translation classroom in a college where translating English into Arabic is the dominant direction for translation seems to be a logical departure point when considering putting previous research

recommendations into research practice. Based on research (Alotaibi 2014) as well as personal observations, MT seems to have been used for quite a while to informally assist native Arab students with their translation tasks. Yet, how effective it is in terms of productivity and quality when students translate from English to Arabic is a question still in need of an answer.

Using MT requires several skills and strategies that are different from those of translation. Among the different skills and strategies associated with MT, PE appears as a 'profession on its own' (Rico Pérez and Torrejón 2012: 166) in which translators use a series of competences such as core competences (which include but not limited to: excellent writing skills, in-depth cultural knowledge, and sound research skills), linguistic skills and instrumental competences to handle MT output in order to gain productivity while maintaining the quality of human translation. Evidence of the need for translators who master PE skills can be seen on the websites of many language service providers in the Arab-speaking world in countries such as Saudi Arabia, Egypt and United Arab Emirates where PE services are offered (Haji Sismat 2016). However, there is a lack of research on PE of English-Arabic MT even though research of PE processes and the opinions, both among experienced translators as well as students, has yielded mixed results in other language pairs (Çetiner 2018; Daems 2016; Doherty and Moorkens 2013; He 2014).

This has motivated me to consider the following questions about MTPE process and product: What if we implemented an educational intervention that included hands-on training on PE of Arabic MT in a classroom where English-Arabic is the language pair in use? Will students gain productivity, or produce more words per day, through MTPE when compared to when they translate manually? Would they be able to maintain the quality of translation when they use MTPE? Are the errors in texts resulting from MTPE and human translation (HT) similar or different? And if they are different, in what way, and what would the effect of these error types be on the designing of MTPE-related translation training course?

Furthermore, research has recommended that for technology users to become actively engaged with the technology and for it to show effective results they must reveal a positive attitude towards it, or acceptance in other words (e.g., Daems 2016). In a definition that relates to the present study, *technology acceptance* involves users finding the technology *useful*, and *easy to use* (Davis 1989). This emphasis on the importance of investigating users' acceptance in the overall translation process has motivated me to include students' opinions about MTPE as an essential part of the comprehensive evaluation of the effectiveness of the MTPE training intervention.

## 1.2 Personal Motivation

There are several issues that motivated me to conduct this research. They relate to the results of previous research, my observation of the performance of translation students, my personal interest in the subject matter as a translation instructor for more than a decade, as well as some implications from previous research that tackles the future training of translators in Saudi Arabia and the Arab World.

Firstly, as previously mentioned, research involving translator training in the Arab World in the last decade highlighted both the gap between the current status of translator training and the requirements of the translation job market there (Al-Jarf 1999; Alenezi 2016; Abu-ghararah 2017) as well as the errors in Arabic MT systems that require human intervention (e.g., Khresheh and Almaaytah 2018). Technology-oriented translator training kept appearing as a promising means to bridge that gap if it was introduced into the curricula (e.g. Abu-ghararah 2017). However, because of the lack of technology-oriented training in academic programs, current translation graduates are not preferred for hiring in organizations that request high-quality translation at a fast pace, or list *CAT skills* as a requirement in their job specifications due to the graduates' lack of knowledge and experience in this area (Abu-ghararah 2017; Fatani 2009). Instead, technology-oriented bilinguals are preferred for the jobs (Fatani 2009).

Secondly, my personal teaching observations played a major role in defining the research questions. Due to the fact that in my college, CAT is offered as a single 2-hour per week module that mainly offers theoretical knowledge with little practical engagement, students who had been using MT systems started utilizing these systems based on personal or zero training. Despite the ethical regulations, I have noticed that students in specialized translation courses (such as legal translation, political translation, and translation of the media) were using the available MT systems more frequently to complete their assignments to benefit from their speedy results and their virtual zero cost, without knowing how to repair the compromised quality of the MT output. In addition, I have noticed that the translated texts (TTs) that graduating students presented during our meetings in the course titled *Graduation Project* (in which every student is required to translate 15000 words and present a written document of the translation) were usually characterized as clumsy, inappropriate, or literal renderings of the source text which indicated to me that students might have been, again, utilizing MT systems without knowing how to fix even the smallest errors in the Arabic MT output, such as, punctuation marks.

Thirdly, my personal interest has also played a role in shaping the research idea and provided me with the momentum to maintain my motivation throughout. I have worked as a supervisor of the course *graduation project* for over three academic years and an instructor of specialized

translation courses for over six years, including teaching the previously mentioned CAT module for four consecutive semesters. This work experience has urged me to investigate technology-related translation methods to find possible ways to improve the methods of teaching and the contents of undergraduate translation courses in my college. Thus, after reviewing the literature in which Arabic MT errors were identified and human intervention to fix them was recommended, my overwhelming motivation was the lack of research concerning the effectiveness of MTPE training, which has become a common practice in global translator training. My personal desire was to explore this issue in my undergraduate English-Arabic translation classroom. This gap in the literature has highly motivated me to convert the idea into a research proposal that will contribute to the literature of translator training especially with regards to Arabic MT training in the undergraduate classroom. So, I prepared a proposal that involves the CAT classroom in a Saudi university, which I am very familiar with and experienced in, where there is a lack of fit between the needs of students which have been identified through previous research on MT training, and the needs of the job market. Further readings concerning students' needs and attitudes towards technology in Saudi translation colleges, and global research on MT and PE training have helped me shape my PhD research proposal (e.g., Al-Jarf 2017; Alotaibi 2014; Gabr 2001; Koponen 2016; Krings 2001; O'Brien 2002).

Lastly, it is worth mentioning that the nature of the current study, which involves running an evaluation of a translation course intervention that involves MTPE, can be considered an application of the recommendation given by Alenezi who suggested that due to the fact that 'there has been a lack of research on translator training in general and in Saudi Arabia in particular, which motivates researchers to explore this area.', that there is 'a chance to conduct research on study materials used in translation classes which this study has generally touched upon. Having the highest percentage of students neutral not only towards their satisfaction with study materials, but other areas too, signals a need for investigation of that area.' (2016: 363)

This research is funded by the Saudi Cultural Bureau in London on behalf of the College of languages and translation (COLT), King Saud University, situated in Riyadh, Saudi Arabia. COLT is a government public college specialized in translation teaching and training. It is the college where this research is taking place, and it is divided into two departments: women and men. Due to the gender-segregated nature of education in Saudi Arabia (Baki 2004), this study is taking place in the women's department where I have been working since 2004. According to the Women's English Translation Department website, the program's mission is '[t]o provide distinguished academic education in the fields of languages and translation, and production based on the economics of knowledge innovation and creativity and contribute to the service of society' ( كلية اللغات والترجمة 2019). The academic objectives of the English Translation department in COLT are:

(1) Preparing specialized translators in modern languages, (2) developing the skills of students in the field of language learning and training in the translation work of all kinds, and (3) preparing, designing, implementing, and reviewing of the new language programmes and translation as needed and in accordance with the market needs and within the University regulations ( كلية اللغات والترجمة 2019). These listed objectives make a solid basis for the main aim of the current study. In fact, in the hopes that the findings of this research will aid in the preparation of technology-oriented and experienced translation graduates who meet the needs of the translation industry, this study will attempt to evaluate the potential effectiveness of a training programme that involves PE guidelines taught to undergraduate translation students which aims to develop translation instrumental competence of those students, particularly when dealing with MT.

## 1.3 The Aim of the Study

The aim of this study is to gain a better understanding of MTPE training in a female undergraduate translation programme in Saudi Arabia by comparing its process and product with an established practice in the translation classroom, i.e., HT. The overall purpose is twofold: that the data and findings of the research provide information that will help decision makers (administrators, translation teachers and students) to judge the merits and worth of implementing MTPE training in undergraduate translation programmes and help developers of Arabic MT focus on the areas in need of constant improvements. In addition, it is hoped that the research adds findings drawn from an evaluative study of MTPE training that will emphasise the importance of incorporating translation technologies in translation training in the Arab World (Abu-Ghararah 2017; Alenezi 2016; Alotaibi 2014; Fatani 2007; Fatani 2009).

The effectiveness of MTPE is evaluated through two aspects of the translator training: the process in which students perform the translation task (translation productivity) and the outcome of the performance (i.e., the quality of the final product). In addition, it is believed that translators' opinions about MTPE matter. They should not only gain faster results while maintaining the translation quality of HT but also feel happy about that (Daems 2016). Therefore, investigating students' opinions would also help provide a holistic evaluation of the effectiveness of MTPE in the translation classroom.

In the present study, the first research question (RQ1) will explore the opinions of the students, the second research question (RQ2) will investigate the effect of the independent variable (PE training) on the first dependent variable (productivity) through a comparison with HT, and finally, the third research question (RQ3) will investigate the effect of the independent variable (PE training) on the second dependent variable (translation quality) in a process of comparison

between HT and MTPE. Thus, an evaluation model that incorporates these three aspects of translator training is needed as well as one that can accurately locate which aspect of the training course was effective or not effective. Also, the design of the evaluation model should follow a logical sequence so that if the training programme does not meet its objectives, it is easier to find out what went wrong and in which level of the evaluation process. It is for this reason that the Kirkpatrick Model of Learning Evaluation (Kirkpatrick and Kirkpatrick 2016) was used.

Thus, in order to answer the main RQ of the study (How effective is MTPE training in a female undergraduate translation programme in Saudi Arabia?), I intend to investigate the opinions, the productivity, and the quality through answering the following RQs:

**Opinions:**

**RQ1:** What are the differences in students' opinions about HT and MTPE?

1.1. How rewarding is MTPE compared to HT?
1.2. How useful is MT output according to translation students?
1.3. Which translation method is perceived as being faster?
1.4. How is the quality of both methods of translation perceived?
1.5. Which translation method is the most preferred?
1.6. Is there a difference in students' opinions before and after the intervention?

Research on opinions up to now is limited and has yielded mixed results. To gain better insights into translation students' opinions about MT and PE, I will conduct focus group discussions before the PE intervention takes place, then use retrospective pre-test surveys right after students have participated in the PE task.

**Productivity:**

**RQ2:** What are the differences in the process between HT and MTPE?

2.1. Is MTPE faster than HT?

In order to gain insight about the productivity of the translation and PE process, I need to use a tool to measure the students' productivity. Students' records of total translation time will be used to register the task total period in minutes.

**Quality:**

**RQ3:** What are the differences in the product between HT and MTPE?

3.1. What are the most common errors in HT and MTPE tasks in the language pair English-Arabic?
3.2. Is there a difference in the overall quality between the product of HT and the product of MTPE output?

To identify error types in translations, evaluators' annotations will be used to calculate the number of times (mentions) each error type was annotated. Also, to evaluate the quality aspects, I need a translation quality assessment approach suitable for a comparative analysis of error count in HT and MTPE texts.

## 1.4 Contextual background of the study

The Kingdom of Saudi Arabia is located in the Middle East, southwest of Asia, and has a population of 34.3 million people as of 2018. Arabic is the official language in Saudi Arabia while English is considered the official foreign language for all sorts of communication with foreigners. In addition, modernization and globalization has played an important role in the flourishing business of translation from English in Saudi Arabia.

In 1955, the first documented translation project in Saudi Arabia (which is a national development project that targeted educating people about different aspects of human knowledge, including science and literature) was launched (Al-Otaibi 2015). That is a phase when Saudi Arabia attempted to fulfil its political agenda through translating into Arabic to educate its citizens through translated books, usually from English and French, in medicine, sciences, literature and humanities, and translating from Arabic to spread the religion of Islam through translations of Islamic Heritage into other languages (Al-Otaibi 2015).

However, since the year 2000, Saudi Arabia has witnessed a shift in translation interests when the main purpose for translation training was to cope with fast spreading globalization and localization, and to keep up with the multiplying needs of the translation industry (Abu-ghararah 2017). The Arab world started hosting translation programmes in the 1960s (namely in Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Sudan, Syria, Tunisia, UAE, and Yemen). Of which, I counted ten educational institutes in Saudi Arabia that offer BA (and several MA) in translation.

Translation training in Saudi Arabia is still under-researched, particularly research on the use and effectiveness of translation technologies. Perhaps, the main possible reason is that teachers as well as students still consider CAT tools to be of secondary importance and that HT is the ultimate objective of translation training. Fatani (2009) highlighted this issue by noting that in many cases the Saudi translator still prefers the conventional approach to translation, i.e., searching for terms in the dictionary even though recommendations for active steps to be taken in order to embrace Arabic MT projects were proposed over 30 years ago (Sieny 1989), and again a decade later (Homiedan 2001).

Conducting the research in a translation college in Saudi Arabia was chosen for a number of reasons: (1) for the fact that the under-researched language pair, i.e., English-Arabic is an official language direction in this college, however, the conclusions of the study might be able to influence pedagogy in translation programmes beyond this one (translation programmes in the Arab World and perhaps different parts of the world where this language pair is offered for training), (2) for practicality and ease of access to participants: as I have worked for the past 17 years in the college where I collected the data, it was easier for me to have access to the students in that specific college, and lastly, (3) for the findings of the study would be of personal benefit for my career as well as the educational institute where I work (which are considered essential factors to keep me motivated throughout the research).

## 1.5 The Socio-technical context for MTPE

The current study focuses on examining ways in which technology works to improve the welfare of translators. Hence, describing the socio-technical context of the study is key.

In the Arab World as well as the rest of the world, globalization has been the spark of the use of MT (Allen 2003). The need to cope with the huge amounts of translation jobs has encouraged many corporations to shift from being local to being multinational. Also, expectations regarding the quality of the translated material have enormously affected the industry. Although globalization and the need for better quality translations justified the urge to improve the quality of MT output, nevertheless, Arab universities and governmental organizations practically showed no interest in funding research or working on projects to improve MT output (Alanazi 2019). In fact, newly developed MT systems were not popular in the Arab World among governmental and private sectors back in 2015 (Almutawa & Izwaini 2015), and the linguistic-oriented approach to translation training added to the issue (Thawabteh 2013). The tendency of governmental and private sectors not to implement MT systems in their work can be linked to the unsuccessful output of automated translation which decreased their expectations of the role that MT systems can play in the industry (Alotaibi 2014). According to the specifications of MTPE, fixing the quality of raw MT output and gaining productivity to cope with the increasing demand are the main pros of MTPE.

In Saudi Arabia, where this study is taking place, the translation profession has been characterised as hard work, low income, bosses' pressure and tight deadlines (Alotaibi 2015). If

training and implementation are carried out professionally, MTPE could tackle at least two of those raised issues about the translation profession in Saudi Arabia, namely hard work and tight deadlines. This is being said because once translators learn how to PE raw MT outputs, a major part of the translation effort is lifted leaving the translator with the mission of fixing the output to publishable quality (Vieira 2019). However, as Al-Jarf (2017) explains, the translation teaching methods currently used in Saudi Arabia are outdated and traditional where students are presented with theoretical aspects of translation studies. These methods give students texts to translate without introducing them to latest translation technologies such as MT systems nor do they offer means to tackle raw MT output.

This study examines MTPE as a method which tackles the issue raised by Alanazi (2019) regarding the lack of involvement of universities in the Arab World through presenting a training intervention which can be used to train students to gain more confidence and experience when dealing with MT systems and hopefully eliminate the hardships of hard work and tight deadlines which were raised by Alotaibi (2015).

## 1.6 Organization of Thesis

This thesis comprises six chapters, including this introductory chapter.

**Chapter 1** provides the rationale for the study, the personal motivation, the aim of the study, the contextual background of the study, and the organization of the thesis.

**Chapter 2** provides general literature and the theoretical aspects related to this current study, such as Holmes' Translation Studies, translation pedagogy, the conceptual framework, and the key concepts involved in this study.

**Chapter 3** describes the methodological approach adopted in this study, involving the clarification of the pragmatic approach, experimental and mixed-method design. It also provides a general overview of the relationship between the research questions and methods used in this study.

**Chapter 4** presents the findings and results from the data collected throughout the study. It begins with the qualitative findings from the Focus Group Discussions held before the teaching intervention to explore students' opinions about HT and MT then quantitative results from retrospective pre-test surveys, and the result from HT vs. MTPE experiment follow.

**Chapter 5** offers the discussion of the findings and results of the study and a conclusion of specific answers to each research question in the form of an outcome of evaluation using Kirkpatrick Model of Learning Evaluation.

**Chapter 6** concludes the study. It starts with the thesis overview, addressing the research questions, and then it presents the contributions of the study, the limitation and difficulties that I experienced during the PhD journey. The implications and the recommendation for further research are also provided in the last chapter of this research.

## 1.7 Chapter Summary

This chapter has offered an introduction to the study. It started with a description of what the study entails. Following, it offered the rationale for the study then the personal motivation. The study aim has been outlined and the purposes of the study and the research questions have been addressed. The contextual background of the study has also been described. Finally, this chapter provided an overview of the organization of the thesis that includes all the chapters. In the next chapter, I provide a discussion of the literature review and the conceptual framework that underpins this study.

# Chapter 2 Literature Review

## 2.1 Introduction

This chapter discusses the academic literature that is relevant to the main aspects of the current research. First, it gives an overview of the current status of translation training in the Western world, comparing it to translation training in the Arab world (Section 2.2.1). Second, it attempts to locate the current research project in the wider field of research by demonstrating its location within the 'translation studies' diagram proposed by Holmes (Section 2.2.2). Third, the conceptual framework will provide the key concepts and structure of the model used in the current study for the evaluation of teaching intervention (Section 2.3). Finally, in an attempt to demonstrate the gaps in research which are addressed by the RQs of the current study, this chapter reviews the existing empirical research into MTPE which involves opinions and productivity as well as quality (Section 2.4). Section 2.4 of this literature review examines research into opinions about MT and PE, in which findings have been drawn from surveys as well as experiments. This section also looks at previous studies that identify differences in translation productivity between HT and MTPE. Furthermore, the research on translation quality is reviewed, especially with regard to translation quality assessment (TQA) approaches that have been used in studies, including those which compare the differences between HT and MTPE. This chapter also review studies that have identified errors in Arabic translation and Arabic MT systems, in order to provide a reference for comparison with our later results. The chapter concludes with Section 2.5, which provides a summary and concluding remarks.

## 2.2 Background

### 2.2.1 The Current Status of Translation Training

The impact of globalisation as well as the increasing requirements of modern life comprise consistent challenges for pedagogy in general, because students need to learn new skills as well as practise autonomous learning in order to cope with those requirements. Translation pedagogy is no exception to this. Without a doubt, globalisation has caused a shift in translation pedagogy in many parts of the world in order to cope with the ever-increasing need for translations for both localisation and dissemination purposes. One of the significant concepts within this pedagogical framework is the inclusion of new technologies in translation training (i.e., technological or instrumental competence). It has been suggested that in order to achieve effective translation training that meets the growing needs of the translation industry, teachers are expected to

introduce changes into the curricula and to design training courses that benefit from current advances in translation technology, while demonstrating its limitations (Varela-Salinas 2020). The suggested equation for success was that if a technology that is integrated into training is considered rewarding, or when students acknowledge its value in their translation tasks, while productivity is increased and quality is maintained, then translation training is considered effective (Daems 2016). Translation programmes in Australia, China, Europe, the USA, and other parts of the world have taken measures to ensure that their graduates fulfil the needs of the translation market (e.g., Çetiner 2018; He 2016; Koponen 2016; Moorkens 2018; O'Brien 2002). MTPE has been highlighted as one of the methods of preparing future translators that are better suited to the job. The importance of MTPE has been highlighted throughout the steady increase in teaching it and developing its guidelines (e.g., Haji Sismat 2016; Kenny and Doherty 2014; Mellinger 2017; Moorkens 2018; Rossi 2017).

Translator training in many countries around the world has largely benefited from developments in the field of translation studies, foreign language teaching, and linguistics, which have brought about systematic approaches to the teaching of translation. New debates about incorporating technological skill sets, learning outcomes, new pedagogical considerations, and curricular content have begun to dominate discussions in the field (e.g., Eser 2015; Ivanova 2016). Findings from such debates have led translation programmes around the world to embrace translator training that moves away from the dilemma over theory versus practice and towards one that moves into teaching models where knowledge is constructed by exploiting technology to fulfil market needs rather than borrowing and copying training programmes that might not suit the local translation industry. These findings have paved the way for welcoming the integration of the post-positivist model of Kiraly and Hofmann, which calls for translator training:

> [T]o focus both on the actual day-to-day practice of work placements, as well as on theoretical considerations that might help justify and explain the incorporation of work placements into translation studies curricula. (2016: 2)

A comprehensive application of this model of translation training was presented by Varela-Salinas (2020), who introduced technological competence into the translation training classroom for the first time in 2007 as a result of acknowledging the importance of technology as an essential tool in the mastery of translation work. The author presented examples drawn from her personal experience in translation training (academic years 2015/2016, 2016/2017 and 2017/2018). The purpose of those proposed courses was to teach linguistics and information communication technology (ICT) skills in addition to professional habits to Spanish students who had studied German translation for three years. The proposed course was introduced after the students had completed learning general translation skills and as they were being introduced to specialised translation. Technology-based training was evaluated through the teaching of three units: MTPE;

the use of corpus and terminology management programmes; and translating with a translation memory system (TMs). The study concluded that in the case of poor linguistic knowledge, CAT tools can compensate to a certain extent, and that students who master a certain level in both the source language (SL) and the target language (TL) are the ones who benefit the most from exploiting technology. It would be interesting to compare conclusions drawn from exploiting technology in a different setting (i.e., the English-Arabic translation classroom in an Arab country).

Recommendations for including technological competence in translator training in the Arabic-speaking world have appeared in a number of studies. For example, Haji Sismat has suggested that 'there is a need for translator training so the translators can utilise the technologies to its [sic] maximum potential and explore which one of them is more suitable for them to enhance their productivity' (2016: 24). Furthermore, Albatineh and Bilali (2017) compared the profiles of translators who are shaped by translation training programmes in the Middle East and North Africa (MENA) to the expectations of the regional translation markets there. In order to systematically run the comparison, the authors analysed the content of courses cited in translators' profiles and compared it with the requirements of translation job advertisements on the market. The contents of training courses were extracted from 61 active translation programmes (21 bachelor's degree and 40 master's degree programmes) in the following countries: Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Morocco, Palestine, Qatar, Oman, Saudi Arabia, Sudan, Syria, Tunisia, Yemen and the UAE. They analysed fifty translation job advertisements in total that were gathered over a four-month period. Forty-six advertisements were seeking full-time translators, while four advertisements were seeking freelancers. The findings revealed that technological skills are extremely in demand in the translation market in MENA, as they appeared in 50% of the analysed advertisements. However, the findings also revealed that academic training did not place similar emphasis on learning technological skills. This finding was evident in the fact that only 22% of the surveyed academic programmes offered courses to develop technological skills.

Before moving on, it is important to make the distinction that, within translation pedagogy, the scope of the current project tackles translation 'training' and not translation 'education'. Translator training is a subdivision of translation pedagogy, which constitutes a relatively more recent sub-field of translation studies, as the highest volumes of articles and research on translation pedagogy started to appear around the last two decades of the 20th century (Piotrowska and Tyupa 2014). Malmkjær (2004) provides a clear distinction between translation 'training' and 'education'. First, the purpose of 'training' is to prepare students to solve problems that they can identify in advance through the application of pre-set procedures, while in 'education', the aim is the overall growth of the individual and the development of their cognitive

capabilities and attitudes that will place them in a position to cope with the different situations within a profession. Second, learning through training is cumulative, i.e., the learner is required to put together as large an inventory of pieces of knowledge as possible in the field in which they are being trained, while in education, learning is generative where the aim is to develop the ability to employ available knowledge to solve arising problems. This is especially important because technology is constantly evolving, and translators need the skills to master technology as it develops over the course of their career, and they may need to do so independently, as they will no longer be in education. Finally, the 'training' approach is useful for teaching translation in specific fields/topics while the 'translation education' approach is more focused on general purposes (Malmkjær 2004). In short, according to the distinction made by Malmkjær (2004), it can be said that 'training' provides students with the skills required to ultimately achieve a level of translation 'education' that it is hoped will future-proof them.

The current study intends to take steps towards closing the gap between English-Arabic translation training programmes and the translation industry that makes use of this language pair by including a learning unit that assesses MT and PE and compares the results with studies conducted in other parts of the world. It also incorporates Kiraly and Hofmann's recommendation to focus on actual day-to-day practice, and considers the technological skills and knowledge components suggested in the ever-evolving translation profession which need to be incorporated in the training curricula (e.g., Al-Jarf 2017). This is because, like the rest of the world, the translation market in the Arab world defines the profiles and professional goals of the offered jobs, while academic institutions are required to develop courses to meet those job requirements (Al-Batineh and Bilali 2017).

### 2.2.2 The Location of the Current Study within Translation Studies

The aim of this section is to illustrate the areas of translation studies that are addressed by the current study. By doing so, it is hoped that the significance of this study is clarified. The illustration below is drawn from the *Pure* aspect of Holmes' Translation Studies map, and Munday's extension of the *Applied* aspect.

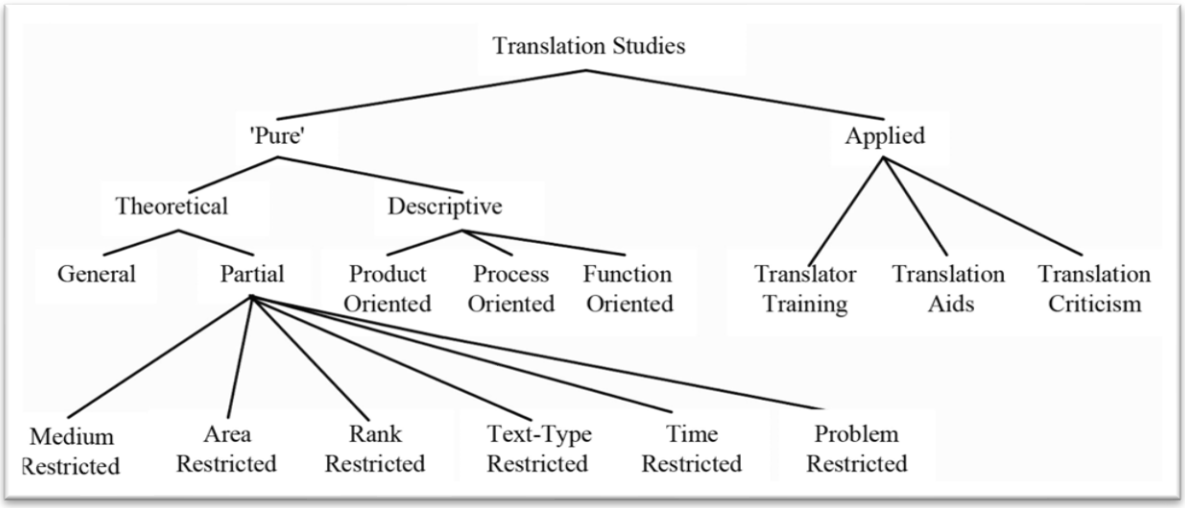
Holmes's 1972 attempt to create a Translation Studies map (Munday 2016) has been widely accepted as a seminal piece in the development of the field as a distinct discipline. For instance, Gentzler described Holmes's paper as 'generally accepted as the founding statement for the field' (2001: 93). However, Holmes's map was subsequently criticised for not including all the different disciplines in the field (e.g., Pym 2017). Holmes attempted to describe what translation studies covers by putting forward an overall framework that divides translation studies into *pure* and

*applied* (see Figure 2-1 below). However, the categorisation within Holmes's diagram of translation studies is not clear-cut. Munday implied that a single project can tackle concepts from both branches of the map:

> The divisions in the 'map' as a whole are in many ways artificial, and Holmes himself points out that the theoretical, descriptive and applied areas do influence one another. (2016: 19)

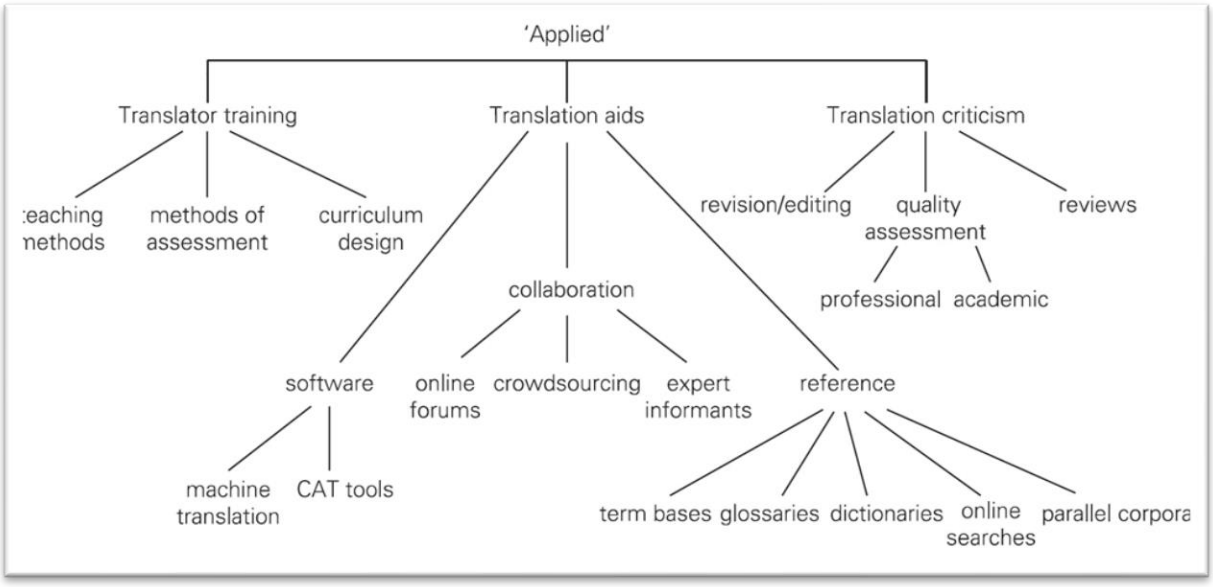
For instance, the current research study can be categorised as one that addresses both the *Pure* aspect as well as the *Applied* in the sense that it is concerned with evaluating the effectiveness of the process (translation productivity) as well as the product (translation quality), which correspond well to the *Pure- Descriptive- Process-oriented and Product-oriented* aspects on Holmes's map. It also tackles the *Applied* aspect, which includes *Translator Training, Translation Aids, and Translation Criticism.* However, according to Munday, 'Holmes devoted two-thirds of his attention to the "pure" aspects of theory and description' (2016: 20) rather than to the applied division of translation studies.

Figure 2-1 Holmes's conception of Translation Studies (Munday 2016: 17)




Therefore, Munday (2016) attempted to expand Holmes's classification of the *Applied* translation studies (see Figure 2-2). Within Munday's conception of Translation Studies, the current study will add specific findings to three of the classifications under the *Applied* aspect of the map: (1) 'teaching methods' of 'Translator training', (2) 'machine translation' of 'Translation aids', and (3) 'academic- quality assessment' of 'Translation criticism'.

Figure 2-2 Munday's conception of 'Applied Translation Studies' (2016: 20)




First, the overall study tackles the issue of *translation teaching methods* that are intended to close the gap between the needs of the translation market and the profiles of translation graduates (e.g., Abu-ghararah 2017). Second, the current research takes into consideration the recommendations of adopting translation technologies raised by several researchers (e.g., Abu-ghararah 2017; Albatineh and Bilali 2017; Alenezi 2015). By doing so, the current study will add to the literature on *machine translation software*, under the general umbrella of Munday's Translation aids. Finally, the current study intends to add to the literature on *academic assessment of quality* that lies under *translation criticism*. This is carried out through the evaluation of student TTs presented before and after taking the teaching intervention (MTPE course).

### 2.2.3 Arabic MT: From Statistical MT to Neural MT

MT was one of the results of the technological development in the twentieth century. As for the interest in Arabic MT, Elsherif and Soomro (2017) claim that it began in the 1970s, whereas Yngve (2000) mentions that Arabic was one of the languages that he worked on in the COMIT project at MIT in the late 1950s. However, it was not until 1990 when research on statistical machine translation (SMT) began (Soudi et al. 2012). Ameur et al. (2020) provide a full summary of research conducted on Arabic MT in which they state that research studies in Arabic MT are mainly devoted to SMT and NMT, while other translation approaches receive less attention.

In SMT, rather than relying on grammatical rules, the computer learns the target language through probability theory (Alqudsi et al. 2014). This approach to automated translation requires

enormous training data. According to Systran (2018), SMT can produce satisfactory language fluency in addition to its ability to catch exceptions to linguistic rules. After rule-based machine translation and example—based machine translation, the statistical approach has introduced a major contribution to the field of MT because of the increased interest in its output (Alqudsi et al. 2014). However, the tendency of overgeneralising persists as a weakness in SMT during text processing, which may lead to inaccurate translation (Cavalli- Sforza & Philips 2012).

However, in recent years, research in MT has switched from the traditional SMT to NMT (Ive 2017), and Arabic MT is no exception. Deep neural networks were first introduced to the translation field by Kalchbrenner and Blunsom (2013) and Sutskever et al. (2014), resulting in the more recent automated translation approach known as NMT. NMT is similar to SMT in the fact that large data collections are needed for it to perform better, however, a major difference stems from the fact that NMT is heuristic in nature, i.e., it learns without the need for human intervention (Ameur et al. 2020). Nevertheless, Berrichi and Mazroui (2021) list a number of challenges in Arabic NMT. For example, they point out that despite the promising quality resulting from NMT in language pairs of similar structure, the different structures of English and Arabic have negatively affected the performance of English-Arabic NMT. Also, because of the limited vocabulary size that NMT models require, the vocabulary coverage rate of Arabic, which is known for its rich morphology, has been decreased. Similarly, long sentences, which are considered a feature of the Arabic language, present more challenges to NMT, which performs less well with longer sentences.

The main focus of research work has been different between SMT and NMT. Ameur et al. (2020). Elaborated on the differences between research work in SMT and NMT: In SMT, improving the quality of SMT has been the major concern for most of researchers working in the field, whereas pre- and post- processing has been the major focus in NMT research. Unfortunately, the main focus of research conducted between the language pair Arabic and English has been focusing on Arabic into English. In addition, the majority of studies that examined English-Arabic MTPE has been conducted using SMT. This limits my ability to run valid comparisons between the results of the current study (being one which used a NMT) and results of previous studies as most of them used SMT systems.

## 2.3 Conceptual Framework

This section identifies and summarises the key concepts relevant to the current research study, including the concepts of effectiveness, opinion, productivity, and quality. The section concludes with a discussion of the learning evaluation model used in the current study.

### 2.3.1 Key Concepts

The next subsections expand on the key concepts that are explored in the current study.

### 2.3.2 Effectiveness

The first question that every researcher asks when commencing an evaluation of effectiveness study is *how is effectiveness defined in my particular study?* This is because there are many definitions of the concept of effectiveness, based on the perspective of the evaluator. For example, are we evaluating a field, a product, or an outcome? In all cases, the evaluation of effectiveness is complex owing to its multidimensionality, i.e., whether the evaluation is quantitative or qualitative, and to the viewpoint of the evaluator (Hamilton and Chervany 1981).

Indeed, different definitions of effectiveness suggest different scopes for the evaluation. One such definition associates effectiveness with the capability to deliver: '[T]he degree to which the system delivers what it is supposed to deliver' (Richards et al. 2019: 5). A second definition examines effectiveness from the perspective of application: '[E]ffectiveness is the application of learning content into work practices' (Noesgaard and Ørngreen 2015: 279). However, both of these definitions are outside of the scope of the current study because it seeks to evaluate the effectiveness of an intervention (MTPE course) in both the process and the product of translation generated by undergraduate students.

Therefore, since classroom research is usually focused on measuring effectiveness through achieved objectives, the one definition that fits well is that proposed by Hamilton and Chervany (1981). Rather than giving a rigid statement that defines effectiveness within a pre-set scope, the authors provide a system for the evaluation of effectiveness. In this system, the researcher determines the task and the objectives of the system being evaluated, then develops measures of criteria to assess how well the objectives are being achieved. Therefore, in the current study, effectiveness is determined by the outcome after comparing the performance of the students to the objectives set at the beginning. An additional advantage of this definition is that it provides measurable objectives that fit well with the learning evaluation model that will be used in this study.

Before formulating a customised definition of effectiveness in the current study, it is important to identify each component of the definition. The *task* in the current study is the *MTPE course*. The *objectives* are that the students achieve productivity, maintain quality, and reveal positive opinions. Now that all the components of the customised definition of the effectiveness are present, a clear definition is set out in the following conditional statement: the MTPE course is

considered effective when the overall result, after comparing translations generated by HT and MTPE students, is in favour of MTPE in terms of productivity achieved while maintaining quality and holding positive opinions of MTPE.

**Opinion**

The importance of evaluating opinions about MTPE is a recurring theme in the literature on PE acceptance and performance. Before expanding on the definitions of different views, it is important to point out that although the terms opinions, attitudes, and acceptance each have a range of definitions and applications in research, in this study, these three terms will be used interchangeably. The reason for this variety in terminology is that although these three words refer to a single concept that is explored in the current study, they are all used by the research papers and articles cited here. The current study investigates the concept of students' attitudes towards and opinions about MTPE skills and use. In a clear example, Daems affirms that '[i]n addition to final quality, translators' attitudes also matter. Even if post-editing is found to be faster, without having to compromise on quality, it is still important for translators to feel happy about their performance' (2016: 26). In addition, Kirkpatrick and Kirkpatrick (2016) state that the first component of the learning evaluation process is *reactions*, which covers the range of feelings and opinions to be examined for the evaluation process to be considered whole. The importance of evaluating the effectiveness of learning through examining students' opinions lies in demonstrating students' perceptions, exploring the possible improvements required to make the learning effective, and providing solid justifications for training expenses (Kirkpatrick and Kirkpatrick 2016).

However, it is essential to commence the study by clarifying and operationalizing the concepts of attitudes and acceptance, and stating the difference between them for the purposes of this study. The concept of attitude has always been a part of social psychology research since its beginning in the early 1900s (Albarracin and Shavitt 2018). However, classification of attitudes is essential as they have different subject matters, or *targets*, that differ according to the discipline of study (Albarracin and Shavitt 2018). Some examples of these subject matters are attitudes towards products in marketing, attitudes towards new medicines in health, and user acceptance in technology-related research.

Attitude has been defined numerous times according to the perspectives of the authors. For instance, Breckler and Wiggins provide a generic definition of attitude as 'a person's evaluation of an object or thought' (1992: 72). However, when attitude is perceived as liking or disliking, it is defined as the 'disposition to respond favourably or unfavourably to an object, person, institution or event' (Ajzen 2005: 3). Nevertheless, technology-related research tends to exchange the word

*attitude* with user *acceptance* and *opinions*. For example, Dillon provides a definition of *user acceptance* as 'the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support' (2001: 218). According to this definition, a user's *willingness* indicates that the user has the choice whether to accept or reject the technology. The factors determining this willingness are the at the heart of the present study. Davis (1989) confirms the importance of user acceptance by referring to it as a crucial factor that determines the effectiveness of technology projects.

In an attempt to formulate the factors that determine user acceptance, Davis defines this acceptance as the *perceived usefulness* and *perceived ease of use*, where perceived usefulness is defined as the degree to which users believe that using the system will enhance their performance, and perceived ease of use is defined as the degree to which the user believes that using the system will be free from effort (1989) (see Table 2-1 below).

Table 2-1 User acceptance: perceived usefulness, perceived ease of use, and their conceptualisation (Davis 1989)

| **Attitude Variable** | **Conceptualisation** |
|---|---|
| Perceived usefulness | The degree to which a person believes that using technology would enhance their job performance (Davis 1989: 320) |
| Perceived ease of use | The degree to which a person believes that using technology would be free of effort (Davis 1989: 320) |

The word 'acceptance' will be used in this research to refer to 'the degree to which students find MT and PE useful and/or easy to use'.

The importance of including the evaluation of students' opinions in the current study stems from personal observations that are supported by findings from the literature, as well as Daems's conclusion (2016) about the importance of translation students' acceptance.

**Translation Productivity**

In this era of globalisation, translation productivity is considered one of the most important variables in the translation formula (Bowker 2005). This is because more and more organisations are seeking to sell their products in the global market and these products require translation into the targeted market's language in the shortest possible period of time.

Evaluating the effectiveness of MTPE on translation productivity in the current study has gained its importance from the previously discussed gap between the translation training in the Arab world and the demands of the translation market there. This study suggests that if students were able to produce higher volumes of translation in less time using MTPE, which includes running the text through the MT system and fully editing it to publishable quality, rather than translating by HT, then the MTPE method would be considered effective (see Chapter 5).

Görög (2015) attempted to provide a definition of translation productivity through considering *time* as the only input and *word count* as the only output, resulting in a definition that 'the more words produced in a shorter amount of time, the higher the productivity will be' (Görög 2015: Paragraph 2). However, the author admitted that this definition is too simplistic as it neglects the temporal effort exerted in editing right after the translation is completed. Therefore, the author suggested that a more reliable definition of translation productivity would be to calculate the total number of translated words in a unit of time combined with the number of final edits done in the whole process of translation. For research purposes only, the current study will use Görög's simplified definition to refer to productivity.

**Translation Quality**

In the ever-evolving world where translation is booming, translation quality and translation quality assessment are becoming core topics in translation research. The importance of evaluating the translation quality in the current research arises from the fact that evaluating the final product is one of the most important ways of judging whether the learning process is effective (Kirkpatrick and Kirkpatrick 2016). Thus, this study will evaluate translation quality by using a reliable TQA in order to evaluate the effectiveness of MTPE as a translation method in the classroom. Nevertheless, before choosing a suitable TQA model that serves the purpose of the study, both the definition of quality and the approach taken should be defined.

The literature provides numerous definitions of quality. A definition is needed that serves the purpose of the current study, preferably one that clearly sets out the variables and the criteria for assessment. House (2014) used three criteria to systemise the different TQA approaches: (1) the relationship between the source text (ST) and the translated text (TT), (2) the relationship between the ST (or its features) and how it is perceived by its author, the translator of the ST, and the reader of the TT, and (3) the steps to be followed when distinguishing TTs from other multilingual text production. Based on these criteria, House provided three TQA approaches:

1. The psycho-social approach (mentalist view), which believes in the 'subjective interpretation of the worth of translation' (2014: 242). Some of the most prominent

authors using this approach are Savory (1952), Gadamer (1960), and, more recently, Stolze (2011).

2. The response-based approach (behaviourist view), which advocates more reliable ways to judge translation. This approach is influenced by the work of Nida (1964), who suggested a number of behavioural tests which allow evaluators to formulate *objective* statements about the quality of a translation. Nida suggested that a translation is 'good' when it achieves 'equivalence of response' (1964: 182), i.e., the manner in which readers respond to the TT is similar to the manner in which they respond to the ST.
3. The response-based approach (functionalist or *skopos*-related view), which downplays the ST and considers the translator as a co-author, while acknowledging that *skopos*, or the purpose of the translation, is the most important factor in translation. This approach is heavily influenced by the work of Reiß and Vermeer (1984).

Previous research involving Arabic MT has identified verbatism as one of the issues resulting from the use of MT in translation, suggesting that it cannot be used even for technical translation (Almutawa and Izwaini 2015). This issue is one of the reasons why the current study follows the *skopos* view, in the sense that it uses a TQA model that prioritises the purpose while maintaining other linguistic features of the TT, such as grammar, style, and spelling, to fulfil the *publishable quality* previously set by the guidelines of full PE (see 3.8 about the TQA model used). Hence, this study uses a definition of quality that fits its scope, as it allows for comparison between HT and MTPE methods in order to judge the effectiveness of MTPE in a translation classroom. According to Teixeira, 'Quality is measured as a score given by professional reviewers, who will process all resulting translations according to predefined criteria' (2011: 4).

**Learning Evaluation Models**

In order to organise and evaluate the effectiveness of a training programme, an evaluation model that serves the aim and objectives of that evaluation is required. There are several evaluation models of training effectiveness to choose from. The researcher considered four models before arriving at the one that best suits the aim of the study. The first was developed by Kaufman and Keller (1994), who adapted their model from Kirkpatrick's model of learning evaluation (further explanation about the Kirkpatrick model below). The difference between the two models lies in that Kaufman's theory of evaluation applies five levels (input; process; individuals and small groups; organisation output; societal outcomes), while Kirkpatrick's model applies four levels. Kaufman's model is not suitable for the purposes of the current study because it seeks to evaluate the programme from a pure trainee perspective as well as assessing the impacts on clients and society which may result from the training programme.

The second model of learning evaluation is Anderson's value of learning model (2007) which is a more recent and lesser-known evaluation model than others that aim to address the two challenges of evaluation and value. The evaluation process is composed of three stages: (1) determining current alignment against strategic priorities; (2) using a range of methods to assess and evaluate the contribution; and (3) establishing the most relevant approaches for one's organisation. The value of learning model encourages organisations to evaluate the effectiveness of learning as a whole, but it fails to provide sufficient direction as to how to measure either the effectiveness or the efficiency on an individual or a small group level. To overcome the challenges, Anderson (2007) suggests that his model is combined with Kirkpatrick's' model of learning evaluation.

The third evaluation model is Brinkerhoff's success case method (2003). The theory behind this model is that any initiative, no matter how successful or not, will always include a spectrum of success and failure. The model seeks to unravel the most impactful successes and failures and then, qualitatively, tell the stories behind them backed by evidence. The limitation of this model lies in the fact that it is not designed to help judge the overall success of a training programme but rather it focuses on the most and the least successful stories. To overcome the shortcomings of this model, researchers in the field of evaluation suggested that it is used alongside other models (Downes 2015).

The last model is Kirkpatrick's model of learning evaluation, which was introduced in the 1950s (Kirkpatrick and Kirkpatrick 2016). As well as becoming one of the most widely accepted and influential models, this model has stood the test of critical review (Phillips 2003). The model involves an assessment of four levels that together make a continuum of complexity, i.e., learners' reactions, their learning skills, their behaviour, and finally, their results. According to Reio et al. (2017), the first level examines the learners' reactions to the learning programme (see Figure 2-3 below). Instruments involved in this level are usually comments about the contents, materials, facilities, and delivery methods. According to Kirkpatrick and Kirkpatrick (2016), the first level, or evaluating reactions, is important no matter what the outcome is; positive reactions to learning may encourage learners to attend the programme, while negative reactions may help in modifying the contents of the programme.

The second level concerns evaluating content. It examines what students have learned in the training programme. According to professionals who used the model (Bersin 2005), the merits of a training programme are demonstrated through evidence that the students have acquired skills from the training. Although research does not support the notion that acquired skills always improve knowledge or equate to behavioural changes (Strunk 2000), research has proven that

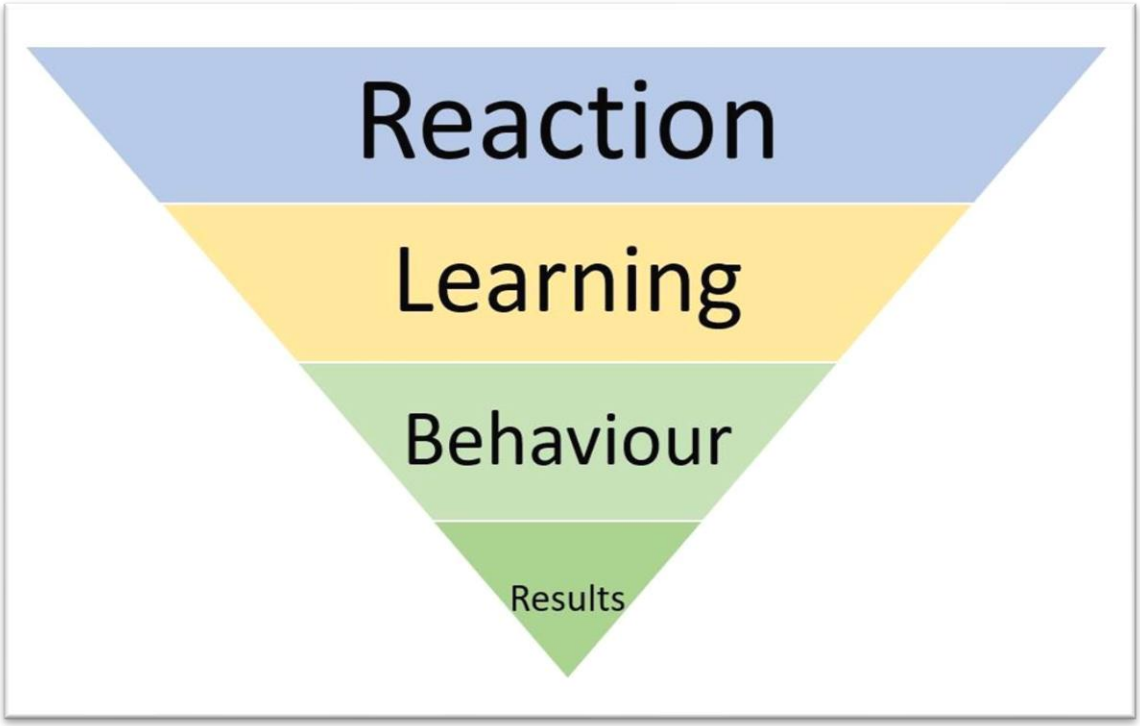
training programme evaluation remains the most important aspect of learning evaluation (Reio et al. 2017).

The third level measures students' behaviour by determining the extent to which the students have applied their new skills to given tasks, since behaviour shows evidence of learning transfer (Reio et al. 2017). The authors note that if students do not apply what they have learned to given tasks, no positive results can be expected out of the training programme.

The last level, or level four, is when organisations attempt to measure the actual results after the training. The pre and post changes are usually measured in numerical and monetary values (Kirkpatrick and Kirkpatrick 2016). All in all, the Kirkpatrick model highlights the importance of examining each of the four levels in order to decide whether they were effective (Kirkpatrick and Kirkpatrick 2016). The advantages of this model are that it can be applied in different pedagogical settings. In addition, its design follows a logical chain, so that if the training programme does not meet its required goal, it is easier for the researcher to find out what went wrong and in which level of the evaluation process (Kirkpatrick and Kirkpatrick 2016).

This model has, however, been repeatedly criticised (e.g., Alliger and Janak 1989). The criticism has emphasised three problematic assumptions in the model as follows. First, there is the assumption that evaluators pay more attention to the higher levels (behaviour and results), while neglecting the lower levels (reaction and learning), due to the hierarchical nature of the model, whereas the model is supposed to be interconnected in order to generate a holistic evaluation (Alliger and Janak 1989). It has been suggested that neglecting the first two levels is a serious mistake that would result in misleading conclusions about the effectiveness of the training programme (Kirkpatrick and Kirkpatrick 2016). The second assumption is that all levels of the model are causally linked; thus, evaluators would presume that positive attitudes must be encouraged so that learning can succeed (Alliger and Janak 1989). The third assumption is that all levels of the model are positively correlated (Alliger and Janak 1989). However, the second and the third assumptions were refuted by Kirkpatrick and Kirkpatrick (2016), who noted that if this were the case, then measuring reactions only could indicate that learning would succeed and therefore performance and results would be positive. The authors emphasised that there are no guarantees that when one level succeeds then so do the other three, and that the uniqueness of the model arises from its structure, since it gives evaluators the chance to identify the levels which requires modification for the learning programme to be considered effective. This model has been used as a powerful evaluation framework in numerous organisational settings, especially classroom training, e-Learning and course evaluation (examples available in Smidt et al. 2009).

Figure 2-3 Kirkpatrick's Model of Learning Evaluation




Considering the strength and limitations of all four different models above, Kirkpatrick's model of learning evaluation would best suit the aim of the present study, while complying with recommendations from the literature regarding the investigation of students' opinions, and, at the same time, examining translation productivity and quality to evaluate the overall effectiveness of a training intervention. Kirkpatrick's model suggests that the evaluation of learning take a holistic approach that is composed of four steps which can be applied in the current study as follows. The first step is to investigate reactions, or *students' opinions* in the current study; the second step is the learning itself, i.e., *the MTPE training*; the third, to evaluate behaviour, i.e., *translation productivity*; and the fourth, to evaluate the results, i.e., *the quality of the translated text* (TT) (see Figure 2.4 below). In Figure 2-4 below, the inverted triangle is the adapted representation in the current study of the Kirkpatrick model, where acceptance (or students' opinions), in blue, corresponds to reactions in Kirkpatrick's' model, training (yellow) corresponds to learning (it is the issue being investigated in the current study), productivity (light green) corresponds to behaviour, and quality (dark green) corresponds to results.

Figure 2-4 Conceptual framework for the evaluation of the effectiveness of MTPE training


## 2.4 Empirical Research on Post-editing

This section starts with the definition of PE and an overview of types. It then reviews studies about PE training and guidelines. The last three sections review the literature on the three RQs of the current research (i.e., opinions about MTPE, translation productivity, and translation quality). The section about translation quality covers the literature relating to the two sub-RQs of the third main RQ (i.e., translation quality assessment and the common translation errors in Arabic HT and in MT).

### 2.4.1 Post-editing Definition and Types

According to Castilho et al. (2018), PE refers to the range of skills that post-editors use to modify raw MT output until the required level of quality is achieved. Allen (2003) classifies PE into two types: *light PE* which is used to correct essential errors and is usually associated with short-life-span texts that are intended for internal use. It is 'strictly minimal editing of texts in order to remove blatant and significant errors and therefore stylistic issues should not be considered' (Depraetere 2010: 2). In contrast, *full PE* aims to achieve publishable quality through running more corrections on different linguistic levels than light PE including stylistic errors. This type of

PE is usually associated with texts intended for dissemination. Around four decades ago, the focus was on light PE. However, even though there have been great advances in MT, the norm for achieving publishable quality is still to fully post-edit the raw MT output. Therefore, it is worth mentioning that in the current educational setting, where students are trained to generate translations of publishable quality, *full post-editing*, or PE that 'implies high quality of the resulting texts' (Depraetere 2010: 2), is the intended skill for training. It is the skill that translation students learn to master in order to achieve publishable quality (TAUS 2013).

According to Castilho et al. (2018), one of the usual concerns in the translation industry is finding a way to quantify the amount of PE effort required to fulfil the final needs and expectations of the end-users. This is important because it predetermines whether MTPE would be time- and cost-effective when compared to translating from scratch. Although quantifying PE effort is out of the scope of the current research, it evaluates the effectiveness of teaching PE skills by comparing error counts in TTs which have been through full PE with those in TTs resulting from HT.

### 2.4.2 Current Status of Post-editing Training

The importance of technology in translation training has been long established through numerous research studies which argue that translation programmes should assist students to become critical users of technological advancements created to help make the translation process faster, more profitable, and more enjoyable, without compromising the quality (Kenny and Doherty 2014). The need to incorporate technological skills in translation training programmes because of growing market demand has been established in many western countries as well as in Turkey, China, and Malaysia. For instance, the results of the survey conducted by Gaspari et al. (2015) suggested that translation technologies are increasingly used in the translation industry. This increased use leads to the strong need for PE training which requires not only linguistic competence but also technological skills. Therefore, researchers who acknowledged the importance of MT and the inevitability of combining it with PE skills have focused their research on MTPE teaching, practices, and guidelines (Haji Sismat 2016; Kenny and Doherty 2014; Mellinger 2017; Moorkens 2018; Rossi 2017). As a result, translation training in many European countries, as well as in the USA, China, and Malaysia, has embraced PE guidelines (e.g., Depraetere 2010; Haji Sismat 2016; Jia et al. 2019; Moorkens 2018; TAUS 2016) in order to prepare trainees for the growing demands of speedy results and tight deadlines, while retaining the translation quality of human translation.

Translation-programme developers in the above-mentioned countries have responded to market demands by actioning recommendations to include MT-related content, PE training, and

technological skills in training programmes. Despite widespread speculation about humanless statistical machine translation (SMT) workflow, Kenny and Doherty (2014) argued that translation educators should include SMT, and they reflected on the role of humans in the SMT workflow. The authors affirmed that rather than marginalizing translators, they should be empowered, and that holistic integration of SMT in training programmes is the way to do it. In addition, PE training and guidelines have become of enormous interest to researchers who recommend embracing technology in the translation classroom. To name a few, Niño (2008) ran an evaluation of MTPE in the foreign classroom and found that PE activities are more suited to advanced learners, while Hu and Cadwell (2016) conducted a comparative study of PE guidelines and found that existing PE guidelines have numerous overlaps, especially for light PE. Koponen (2015) described an MTPE course from a teacher's perspective that focuses on the planning and teaching of the PE course. Similarly, Guerberof Arenas and Moorkens (2019) described the objectives and structure of an MTPE course that was offered in the Localisation Master's programme at Universitat Autònoma de Barcelona in 2009 and in 2017. The search for PE guidelines for the Arabic language yielded only two results: the first in an online Translation Automation User Society (TAUS) webinar titled *TAUS post-editing webinar for Arabic Module* (TAUS Videos 2015), and the second in an unpublished PhD thesis by Haji Sismat (2016), where it is written about from the perspective of non-native post-editors of Arabic.

This inclusion of technological skills in the classroom in such countries has resulted in a level of curriculum evaluation that ensures better preparation for translation graduates. For example, Mellinger (2017) noticed that translation training programmes are increasingly incorporating MT into the curriculum, where the translation graduate's competence in PE skills is assumed. However, rather than the usual stand-alone MT-related training course, which might lead to compartmentalisation of competences, the author suggested that MT-related content be embedded in multiple translation courses. These contents include terminology management, controlled authoring, PE, and engine tuning, which 'involves modifying the MT system and implementing new rules to improve MT output' (Mellinger 2017: 288). This additional material is intended to prepare students for the growing demands of the translation market. Unfortunately, efforts in locating the gap between translation training programmes and market demands in the Arab world seem to have halted at the point of exploring the gap without any research yet available on ways to close that gap.

The likelihood of MT being used in organisations and higher education (HE) institutions in the Arab world has been addressed in a number of studies (e.g., Thawabteh and Territories 2013; Almutawa and Izwaini 2015; Abu-ghararah 2017) For instance, Almutawa and Izwaini (2015) surveyed actual and potential Saudi Arabian users of MT, such as organisations and translation

agencies as well as HE institutions, to investigate how widely MT systems are used and researched in the Arab world. Through a questionnaire, the authors surveyed twenty-one Saudi universities; six of the universities offered a degree in translation, but only three out of those six offered a course in MT. King Saud University, where the current study is taking place, is one of the three universities that teach MT courses; however, it offers only an introductory course in MT to undergraduate students that mainly focuses on theoretical demonstration of MT potentials (Al-Jarf 2017). The survey revealed that despite the increasing interest in MT technology and the growing demand for translation, only 20 of 44 Saudi organisations are using or planning to use MT, and they believe that MT is useful in lexical translation, but not for completing translation jobs. The survey results also revealed that 24 of 44 organisations refused to use MT because they thought that it was not suitable for translating complex sentences, and that MT services are only good for gisting, while publishable translations can only be achieved though HT.

Later, Abu-ghararah (2017) ran a case study about the gap between translation programmes and market needs in Saudi Arabia. She reported a lack of technology and learning resources in the translation programmes of twenty-four public Saudi universities, and found that only eight universities offered translation modules for English language students. This gap between market needs and technological skills was previously identified by Thawabteh and Territories (2013), who noticed that translation programmes in the Arab world are mostly linguistic-oriented at the academic level. Taking these findings into consideration, it would be interesting to evaluate the effectiveness of an MTPE training course offered on a BA translation programme in a Saudi university. Such findings would be beneficial for stakeholders to decide whether or not MTPE training should be adopted in undergraduate translation programmes where English-Arabic translation is taught.

The current study attempts to challenge these findings by comparing HT and MTPE in a classroom setting in order to decide whether MTPE output can achieve more productivity while maintaining quality similar to that of HT. In addition, the training course and PE guidelines that the current study intends to use are partially adopted from a study conducted by Şahin (2014), in which he investigated the potentials of using MTPE activity in the translation curriculum. Şahin created a course in which translation students post-edited three types of texts from different fields (media, technology and law) translated by the Google Translate MT system from English into Turkish. Fifteen fourth-year translation students took part in the five-week study, which involved written logs by the students, students' responses to a post-experiment survey, and the comparison of the final products of HT and PE. The author created a course that focused on MT and PE. The aim of the course was to raise awareness about MT systems and to introduce translation students to the practice of PE. The ultimate goal of the course was to contribute to the development of

translation skills through the use of MTPE. The students attended a three-hour session in the computer lab on a weekly basis for the duration of the course. When post-editing, students were asked to perform full PE to achieve a translation quality that would be similar to HT. The PE course that Şahin used covered the following TAUS guidelines (TAUS 2013):

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.
- Basic rules regarding spelling, punctuation and hyphenation apply.
- Ensure that formatting is correct.

After the conclusion of the final task, Şahin compared the translation quality results of HT and MTPE. The results showed no statistically significant difference between the two tasks. The study was comprehensive in that it drew data from students' written logs as well as their answers to a post-experiment survey, the sample size was small, and the product assessment was all conducted by one reviewer (i.e., the author). Although the current study uses similar course content to evaluate the effectiveness of a post-editing course, in Şahin's study, neither translation productivity nor students' opinions were measured. The current study measures translation productivity and quality for a holistic evaluation of the effectiveness of MTPE in the translation classroom. In addition, the reliability of the results from comparing the two methods of translation is considered in the current study through the assistance of two external evaluators who do not know which method was used for each translated text (see 3.8.1.2 for more about the recruited evaluators).

Yamada (2019) conducted a study to compare error correction rate and effort of SMTPE and NMTPE tasks completed by university students working on the language pair English-Japanese. The results suggested that students' PE of NMT had poorer effort rate despite the fact that their TTs were of higher quality. Although NMT raw output had fewer errors compared to SMT, study results showed no significant differences in students' cognitive effort in both tasks. This suggests that NMT requires more concentration on fewer errors. According to this last result, Yamada (2019) suggested that although NMT is indeed considered a factor that helps produce better-quality TTs, it is unlikely to enhance students' performance while in training. Yamada (2019) also suggested that PE of NMT requires similar translation competence to that of human translation. Hence, training is essential for students to eb able to shift their attention to the right errors and post-edit those errors effectively.

This review of PE training has demonstrated that while advancements in PE training and guidelines are being researched and developed in the West and in countries such as China, PE training in the Arab world and in English-Arabic is yet to be explored and established, although the need for it has long been highlighted.

### 2.4.3 Opinions

Technology is becoming the courier of translation, and one of the features of translation is that it has become a translator-computer interaction (O'Brien 2012). Kiraly (2000) suggests that it is of great importance to try to understand the opinions and perceptions of translators (whether experienced or trainees) towards translation technologies because translators' acceptance is not less important than the final quality of the translation. This emphasises the importance of exploring the acceptance levels of translation students, since they are translators in the making. The following review of empirical research on technology acceptance and opinions about MT and PE highlights studies using students as participants, and those comparing students with experienced translators. A connection between the reviewed studies and the current one will be established by comparing the aims, scopes, and methodologies of previous research with the current study. This review will provide a justification for the current study and considers a number of findings, which will be referred to in the discussion of the results of this study.

In the Western world, translators' opinions about MT and PE have received considerable attention in the field of translation technology (De Almeida 2013). They are considered to be some of the main factors that either promote or hinder the success of implementing MTPE in the daily life of a translator (Doherty and Moorkens, 2013). Acceptance of MT can help translators to embrace and use it to maintain high quality of the final product while increasing productivity (Alotaibi 2014; Çetiner 2018; Daems 2016; De Almeida 2013). However, over time, research has revealed that both professional as well as student translators have shown negative opinions towards MT. They have regarded it variously as a 'job killer' (Krings 2001: 33), as too complicated (Brosnan 2002), or as being slower than human translation (Gaspari et al. 2014). These opinions still exist, especially amongst translation students who are unfamiliar with the potentials and possibilities of MT output and PE skills. However, when the researcher attempted a search for the opinions of female students, or for students' opinions in general, about MTPE in the English-Arabic translation classroom, the search did not return results which provided solid justification for conducting this phase of the study.

Although it was hard to find these studies, there have been some in recent years that have emphasised the importance of examining the acceptance of female technology-users, especially

in an academic context. In one of the earliest studies about technology acceptance in the Arab world**,** Al-Gahtani et al. (2007) examined individual, technological and organisational factors affecting computer acceptance in Saudi Arabia. This large-scale survey used questionnaires that involved 1,190 computer users (238 women) in 56 private and public organisations in Saudi Arabia. Findings concerning women indicated that they had a lower degree of computer usage and satisfaction. The authors linked their reluctance to use technology to the fact that the sample (women in Saudi Arabia) was socially and culturally different to that in a developed country: these women follow certain religious norms and work in environments that separate them from men. Al-Gahtani et al. (2007) suggested that future research should incorporate more women, and that it should aim to shed light on women's computer acceptance and usage to better reflect demographic effects and gender roles in computer acceptance in Saudi Arabia. The present study will explore MTPE acceptance in an all-female translation classroom in order to explore whether they show more acceptance and flexibility when using a CAT tool (MT systems) to produce translations of a similar quality to that produced through HT.

Al Lily (2011) has since echoed the call for more research into women and technology-facilitated communication in Saudi classroom contexts. This would enhance the bottom-up approach, which collects data in the form of responses from students. According to Al Lily, women's use of technology is expected to increase due to the uniqueness of Saudi Arabian society, where women usually study and work separate from men. Women in Saudi Arabia were granted the right to higher education in the 1970s; they study on all-female campuses and are usually taught by female academics or via one-direction CCTV if the teacher is male. In this study, we will address Al Lily's recommendations for further research in an attempt to add to the literature on technology acceptance in translation training in the Arab world.

Indeed, the above-mentioned two studies (Al-Gahtani et al 2007; Al Lily 2011) provide justification for the importance of examining technology acceptance among women in Saudi Arabia. However, the existing literature tends to be general in nature and the field lacks detailed accounts of translation students' acceptance and use of technology, particularly MT technology, in the classroom context. Two recent studies can be considered seminal for the current study because both were conducted at the same college as the current study, and the participants in both fulfilled the same criteria of gender, background, and familiarity with the translation courses at this college. These studies were conducted by Alotaibi (2014) and Al-Jarf (2017). Alotaibi (2014) assessed students' degree of knowledge of CAT tools as well as their expectations and attitudes when using such tools. The study was conducted in the female department at COLT, King Saud University, Riyadh, Saudi Arabia. A total of 103 female Arab participants studying the course *Computer Applications in Translation* in 2011 were asked to complete a questionnaire both at the

beginning and at the end of the semester and to participate in semi-structured interviews. Classroom observations were also used to collect data. At first, many students had relatively high expectations of the quality and the capabilities of some of the CAT tools. They expressed disappointment when some of the tools they tried out returned low quality output and had limited support for Arabic. The results of the study showed that after the students learnt more about the strengths and limitations of CAT tools, their acceptance level increased. Alotaibi suggested that there is a relationship between increased knowledge of CAT tools and the positive change in students' acceptance of technology. According to Alotaibi, 'Students' attitudes became much less biased and, in general, fairly positive' (2014: 65). Alotaibi (2014) concluded by suggesting the integration of technology in general, and CAT tools in particular, into the translation classroom to enhance students' translation skills and maximise their future job opportunities.

There are two potential issues with the methodology used in Alotaibi's study. Firstly, although pre and post questionnaires are widely used in classroom research to measure the effectiveness and impact of educational programmes or interventions, the traditional pre-test-post-test measures assume that the participant's assessment of the measurement will not change from the pre-test to the post-test. According to Lam and Bengo (2003), this assumption may be deceiving because the participant's perception of the construct that is being evaluated may change as a result of the educational intervention, leading to response-shift bias, i.e., underreporting by the participant of any real change that occurred between the pre-test and the post-test. According to Drennan and Hyde, response-shift bias occurs when 'the student's internal frame of reference of the construct being measured is changed' (2008: 699). Therefore, the questionnaires that were used to evaluate change may be problematic due to the confounding factor of response-shift bias. In addition, although no study may proceed without receiving ethical approval, the fact that participants in Alotaibi's study were students that she taught on a mandatory course means that the pre-existing power relationship, which was not taken into account, may have affected the students' responses, especially in the post questionnaire and the interviews. In order to ensure the validity and reliability of results while attempting to eliminate the response-shift bias, the current study uses a 'retrospective pre-test survey' after the PE task. In addition, this research is able to maintain the authenticity of participant responses due to the fact that the researcher has no power over the students' grades, while student participation in the study was completely voluntary.

Al-Jarf (2017) described and assessed the satisfaction and usefulness of CAT training experienced by 72 female graduates who were all working as professional translators/interpreters at the time the study took place. The author used open-ended questions about the two CAT courses offered

by the college, which she analysed qualitatively. Although students are offered two courses on CAT, the findings showed that translation graduates found the course contents inadequate because: (1) the introductory course was taught by teaching assistants, who were unfamiliar with CAT tools, and the course usually offered general lessons about MS Word, PowerPoint, and Excel that were not field-specific; and (2) the advanced course introduced the theory of MT, but students were never given the chance to practice MT or online CAT tools. Therefore, Al-Jarf suggests that if translation programmes in Saudi Arabia seek to prepare graduates for the twenty-first century jobs market, they need to integrate several technologies, such as MT systems, Google Language Tools, and 'how students can efficiently and effectively…improve translation quality.' (2017: 3).

After Al-Jarf's study, research on MT acceptance and use in the translation classroom seems to have halted, and the researcher found only one study that examined the attitudes of college students towards an MT system. This study is important because the MT system under scrutiny was Google Translate (GT). Alhaisoni and Alhaysony (2017) attempted to survey the attitudes of 92 Saudi students who were in their final year of an English programme, towards the use of GT. A questionnaire was used to investigate student attitudes to GT and the reasons for using it. The survey results revealed that the vast majority of students used GT for assistance with learning vocabulary and with reading and writing skills, but not for translation tasks. This might be because these students were learning English but were not trained in translation skills. Nonetheless, the study found that most students had a positive attitude towards GT and considered it to have a positive effect on their language learning because it is free, easy to use, and translates texts quickly. The authors concluded by recommending a framework that addresses the pedagogical implications of GT in Saudi programmes. It is worth noting that although the authors did not define the concept being studied (attitude), the finding that students held positive attitudes towards GT 'use' could link this study to Dillon's (2001) definition of attitude, in which 'usability' is a major factor in positive attitude. Also, although the sample size in this study was quite representative (92 students usually represents a complete cohort in a final year BA in a Saudi university), the authors did not consider the gender of the participants nor the correlation of gender to attitudes towards GT. This is an important point, because previous studies conducted in Saudi Arabia have researched technology acceptance in male rather than female populations (e.g., Al-Gahtani et al. 2007). However, there are no studies which look at the opinions and attitudes of female translation students in similar settings, or which compare male and female attitudes.

Previous reviews clearly show that although research into opinions of MTPE in the Arab world is justifiable and necessary, studies on its acceptance and use have not yet been conducted. This is

not the case in the Western world, where studies have long identified the need to examine users' acceptance towards MT and PE to keep up with the evolving requirements of the translation industry. Accordingly, researchers have conducted studies which focused on the opinions of students' towards MTPE and reported their responses and the recommendations to improve acceptance of MTPE in the classroom.

For example, drawing on data from surveys and reflective journals, Doherty and Moorkens (2013) qualitatively investigated the attitudes of ninety students (final year BA and first-year MA) in an ongoing evaluation of translation technology labs in a university setting in Dublin. The results revealed that most of the students were against MT. The authors were astonished by the findings, given the presumption that '[a]ttitudes are predominantly formed as a result of the student's own direct experiences' (Doherty and Moorkens 2013: 132), and the fact that most of the students who participated in this study had no previous experience of MT and its strategies. The authors attributed this sceptical and biased opinion about MT to be a '*learned* attitude' (p. 132) from others, since very few of the study participants had had any direct experience with MT. Another study-related explanation the authors provided is that students of humanities might find it more challenging to engage with technology than their peers in other disciplines, which eventually limits their ability to exploit MT. The apparent potential rejection of technology could be attributed to students' perception of it as 'not easy' although they eventually acknowledged its potential 'usefulness' in their future workflow. The findings, which were themes developed from the qualitative analysis of reflective journals written every two weeks as an assignment, provided a basis for clear recommendations for MT training in the classroom that included but were not limited to the need to consider the different IT abilities of students and the careful design of lab sessions. It would be interesting to compare the results of the current study with those of Doherty and Moorkens (2013), especially because students have two characteristics in common with the participants in their study: they do not have previous experience of MT, and they all come from a humanities background.

Daems (2016) provides an extensive comparative analysis of HT and MTPE (without training) by both translation students and experienced translators for general text types in the language pair English-Dutch. The author studied the process, the product, and attitudes in detail. The aim of the main PE task in this PhD thesis was to assess the participants' use of keystroke logging and eye-tracking (to examine speed) via a fine-grained translation quality assessment model (to serve as a TQA), and a survey was used to explore the attitudes of participants before and after the PE task. Of particular interest for the present research are the survey statements about attitude suggested by the author. Specifically, the survey contained questions about perceived usefulness and ease of use, perceived speed, perceived quality of MT and PE, and the translation method that was

perceived as least tiring. Daems (2016) detected the change in attitude through administering the survey twice: once before the PE task, and again after it was over. The analysis of attitudes drawn from both students and professional translators indicated that all participants found MT to be useful, however, they preferred HT and thought it was more rewarding. However, Daems did not define what 'rewarding' meant in her work, and neither did Fulford, upon whose study (2002) Daems built part of her work. Participants also believed PE to be as fast as HT, yet only a few participants thought that PE was a better translation method than HT. The current research will involve 'PE training', which Daems (2016: 162) has emphasised as important in improving attitudes, through the use of the word 'understanding' in the quotation below:

> [I]f participants changed their minds after the experiment, it was usually in favour of post-editing, indicating that understanding indeed leads to acceptance.

The study also reported some significant findings regarding productivity and quality. Relevant findings will be covered in the following section dedicated to reviewing literature on translation productivity and quality. Daems's study provides a number of contributions towards research into PE processes, MT training, MT developing, and translation training. The author suggests that the study should be replicated in similar settings but for different language pairs and/or text types. It is worth noting that due to the fact that the author conducted a fairly comprehensive study that employed state-of the art data collection methods, this led to the number of participants being limited in order to for the study to be carried out (23 participants with 13 professional translators and 10 MA students). Although Daems's study employed state of the art data collection methods, the RQs predefine the data collection methods, and Daems's scope focused on evaluating the process and product of translation through measuring PE effort, which lies outside of the scope of the current study. There are also added limitations of the study (one language pair, one text type, and the fairly small sample size that prevents generalisability). Also, the comparison between students and professional translators added extra variables for the task analysis since participants do not share similar profiles (age, background experience, gender…etc) for the findings to be entirely applicable in a pedagogical setting. The current study will attempt to avoid the limitations of Daems's study by using a larger sample size of female students with similar background knowledge, while maintaining the scope of one language pair (English to Arabic) and one text type (technical) for better control of variables.

Lastly, the significance of the following study by Çetiner (2018) is based on a number of similarities between its aim and methodology and those of the current study, as they both explore the opinions of translation undergraduate students. They both examine students' opinions before

and after a training intervention as well as sharing similar sample size. The author statistically analysed attitude questionnaires administered before and after a CAT course that was given to undergraduate Turkish students at Kırıkkale University in Turkey. The study was piloted with 63 students from different departments to test the validity and reliability of the results, then the experimental group of 66 students took a CAT course with hands-on training on CAT tools. The results showed that students in the experimental group revealed positive attitudes, which were interpreted as a tendency and willingness to use CAT tools after the training course was over. According to Çetiner, '[t]his result supports the view that students develop a positive attitude after they are taught the benefits of using computer-aided translation tools and more classes should be allocated for translation technologies in translation training programs.' (2018: 153). Çetiner's study has been criticised for similar methodological and ethical issues as the previously discussed study by Alotaibi (2014), particularly over the potential response-shift bias and the existence of power relations between the researcher and the participants. However, due to the lack of studies on undergraduate translation students' opinions about MT and PE, it would be useful to compare results to find out if translation students translating from English into Arabic reveal similar or different opinions after taking a training course. This would represent an additional similarity with Çetiner's study that has not featured in previous research.

All in all, studies conducted on translation students have yielded mixed results, ranging from positive opinions about MT and its strategies (e.g., Alotaibi 2014; Çetiner 2018) to expressing mixed feelings (Daems 2016), to negative feelings towards it (Doherty and Moorkens, 2013). It can be noticed that the authors of these studies have used quite similar sampling techniques (cluster sampling where homogeneous groups of students participate in the study). However, methods of data collection (questionnaires, student diaries, and pre-post surveys) and analysis (qualitative and statistical) were different which might provide explanations for the differences in findings. Due to the lack of research into PE acceptance in translation classrooms in the Arab world, this study attempts to gain a better understanding of the opinions of translation students there about MT and PE. In order to find out whether their opinions change after a PE training course and a PE task that they perform by themselves, we have used two methods of data collection: focus group discussions (FGDs) before the teaching intervention and translation tasks, and a retrospective pre-test survey that was filled out after the task. A further discussion of data collection methods can be seen in Section 3.9.

### 2.4.4 Translation Productivity

According to Krings**,** 'the time savings that occur (or should occur) in comparison to a pure human translation are accordingly the most important characteristic value in the calculation of the

economic viability of machine translation' (2001: 532). This statement, from a study that is considered seminal for research on measuring PE effort, including temporal effort, has provided justification for numerous studies that measure the speed and/or productivity of PE and run comparisons between HT and MTPE. Such comparisons usually aim at deciding whether MTPE is faster than HT, and if yes, calculating the percentage of time saved between the two methods.

Evaluating the productivity of English-Arabic MTPE in the translation classroom through comparing MTPE and HT is one of the objectives of the current study. A search was therefore made to locate studies that ran comparisons between English-Arabic HT and MTPE, or that examined temporal effort, or measured the speed or productivity of English-Arabic PE. Out of forty-four reviewed studies that tackled PE temporal effort, productivity, or speed, Arabic has appeared in only two (Green et al. 2013; Haji Sismat 2016). Unfortunately, these two studies do not have much in common with the current study. The study by Green et al. (2013) evaluated the efficacy of PE amongst paid professional translators, while the study by Haji Sismat (2016) compared HT and MTPE done by non-native learners of the Arabic language. However, their significance stems from the fact that they both examined a language pair that involved Arabic, which provides the current study with results to compare its results against.

Acknowledging the importance of PE in bridging the gap between MT output and the quality of skilled HT**,** Green et al. (2013) presented a rigorous and controlled analysis of PE to decide whether PE could reduce time while maintaining the quality of skilled HT in three language pairs (English to Arabic, French, and German). In the English-Arabic part of the study, sixteen professional translators were recruited, who were paid for participating in the study. The participants were asked to follow instructions to translate (either through HT or MTPE) 27 source sentences that were extracted from four general texts (two simple and two more challenging texts). The sentences appeared in random order, which ensured that participants did not know the required method of translation they were to use prior to them appearing on screen. Time pressure was applied in the task, but participants were allowed to use bilingual dictionaries. When post-editing, participants were instructed to use GT. They were allowed to submit, manipulate, or even delete the suggested translations from GT. All keyboard, mouse, and browser events were recorded, and the gathered data was statistically analysed. The results suggest that PE reduced translation time while increasing the quality of translation resulting from MT in all three language pairs used in the study (including English-Arabic) with 'very significant effects for Arabic' (Green et al. 2013: 446). Due to scarce studies on English-Arabic PE, the significance of the study by Green et al. for the current research is twofold: its selection of the Arabic language as one of the target languages used in the translation tasks, and its use of GT for the MTPE task. Although the results of the study seem promising, the aim in the present research focuses on finding out whether non-

paid translation students' post-editing of texts generated by GT would present similar productivity.

Haji Sismat's PhD (2016) focused on exploring translation directionality between foreign languages through measuring PE productivity amongst non-native translation trainees (Malay students post-editing Arabic-English-Arabic MT texts) in a translation classroom setting. Six Malay translation students translated from scratch, and post-edited both MT and TM-generated technical texts using MemoQ 2014. The findings revealed that despite the fact that all six participants could not reach the daily productivity of full HT (approximately 2000 words) when they were translating, four participants (67%) managed to reach the daily productivity of MTPE (5000 words) when they post-edited the MT output. Translation technologies improved the non-native translation students' speed by an increased average of 46.2%. Although the present study shares similar classroom settings and level of students (undergraduate students of translation), it is of pedagogical interest to find out whether native translation students post-editing English-Arabic MT generated texts would reach similar productivity as that of the study conducted by Haji Sismat (2016), while maintaining the quality of HT.

As previously mentioned, Green et al. (2013) and Haji Sismat (2016) were the only studies that included English-Arabic MTPE. However, they did not provide a reference for the comparison of the two methods of translation (i.e., HT and MTPE). Since the current study aims to compare HT and MTPE to evaluate the effectiveness of including MTPE courses in translation programmes in the Arab world, studies were searched for in other language pairs that ran the comparison between the two methods of translation. Luckily enough, many researchers have conducted studies that provided such a comparison, but this abundance in the number of studies comparing HT and MTPE required that criteria were set for which studies to include in the literature review. Here, only studies directly relevant to the current research were reviewed, especially those with similar aims and methodology.

Nitzke and Oster (2016) investigated the differences between HT and MTPE through analysing how texts from different domains are processed. The authors used triangulation (total production time, key logging and eye-tracking). In the section of the reviewed study that relates to the present study, nine students took part in post-editing relatively short technical texts (150 words) from English into German using the MT system (CASMACAT). The results confirmed that participants needed less time in PE than when translating from scratch. The present study resembles Nitzke and Oster's in that they both run a comparison between HT and MTPE amongst students, that they both use technical texts for the translation task, and that they both rely on total production time to examine translation productivity. However, we must acknowledge the

existing differences between the studies, which might affect the comparison between the two. First, the reviewed study used a different MT system (CASMACAT) which acts as a different variable. Second, it has a relatively small sample size as it recruited 9 students for the MTPE phase of the study while the current study recruited 60 students in total. The current study takes into consideration the recommendations by Nitzke and Oster (2016) which suggest using different languages pairs (i.e., English-Arabic in the current study) as their study compared HT and MTPE in English-Dutch.

Research on MTPE of technical/scientific texts is already established in many language pairs, but unfortunately not in English-Arabic. Studies conducted on technical and/or scientific texts revealed that PE productivity benefits largely from translators' familiarity with the domain and the nature of the texts. For example, Aranberri et al. (2014) compared PE productivity in technical texts between six professional translators and six lay users who generally did not have experience in MTPE (only two of the professional translators had previously post-edited, and in research experiments only). The lay user group consisted of a group of staff members at the faculty of computer science in the University of the Basque Country who were not specialised in translation but were familiar with the text type that was used in the study. Participants in the two groups were asked to translate English into Basque, through a science and localisation trained MT system that was developed by Bologna Translation Service (BTS). The study involved post-editing two technical texts of around 1,200 words that were rated as moderately difficult, as they address specialised topics. Terminology in both texts was significant; however, the text types were different. The first was mainly descriptive while the second was more literary than descriptive. The authors used the literary text to add some room for creativity which they considered as an added difficulty for the lay users. The findings revealed that productivity increased in both groups by an average of 17.66% (translators) and 12.43% (lay users). The authors suggested that productivity is attitude dependent, as the lay user group seemed to benefit more from the technology because of their previous familiarity with the domain and their willingness to use MT systems, whereas the professional users, who were also familiar with the text types, revealed more negative attitudes as they considered MT a means that would slow down their workflow.

A shared feature between the students participating in the current study and the participants in the reviewed study is that they did not all have previous MT or PE experiences. In addition, both studies shared the same text type (technical). However, the length and difficulty of the texts used in the current study differ because they depended on the criteria set by the college where the current study took place. It would be interesting to compare the translation productivity in HT and MTPE in the current study against that of Aranberri et al. (2014), due to the inexperience of the participants in both studies. Thus, the present study seeks to confirm whether the students'

productivity would similarly benefit from MT when they come to the task equipped with previous familiarity with the domain but without PE experience.

Jia et al. (2019) compared HT and MTPE in the classroom setting to explore the PE process within a neural machine translation (NMT) paradigm. They used domain-specific and general texts in English to Chinese tasks completed by 30 first-year postgraduate students. Similar to the current study, they divided the students into two groups: 15 students in G1 translating from scratch, and 15 students in G2 post-editing MT output. Participants were selected to join either G1 or G2 based on their results in the most recent translation test, to ensure that the level of their translation ability was comparable between the two groups. Domain-specific and general texts ranging from 142 to 156 words were used in the study. The domain-specific texts were a patient information leaflet and a dishwasher manual. Participants were allowed to use dictionaries. They were asked to provide publishable quality through both methods of translation (HT and MTPE). Each of the texts was translated by one group and post-edited by the other group. When analysing the data, the authors calculated the translation speed by dividing the total processing time of each sentence by the total number of words in the ST sentence. They found that MTPE was faster than HT, especially in domain-specific texts. Although the methodology used in their study is different from the one used in the current study, in that participants in the reviewed study did not take a unified pre-test to be used as a benchmark against which post-test changes in productivity could be compared (which is part of the methodology of the current study), it would be interesting to compare the results of the current study to theirs, especially as their participants similarly considered the translation task a lexical one and therefore tended to use the dictionary to search the meanings of unfamiliar terms.

In a similar sense, Yang et al. (2020) presented a report on a productivity test of MTPE. The findings are significant for the current study because the overall objectives of both studies are similar in that they assess the productivity, quality and attitudes of MT users without prior PE experience, and in that they have similar educational settings. Yang et al. investigated the usability of MT in the English-Chinese translation classroom by comparing translation students' MTPE with their HT in two comparable translation tasks. They empirically measured three dimensions of usability: efficiency through measuring the duration and speed of translation, effectiveness, or translation quality, through measuring the number of type of errors found in translations, and satisfaction through a small user-satisfaction questionnaire. A total of 31 first-year postgraduate native Chinese students participated in this study. The study involved tracking students' keyboard activity logging through Translog-II, which records typed and deleted letters, the duration and pauses of the task, cut/paste activities, and mouse activities. The participants

were asked to HT a text and PE another text. The texts were carefully prepared so that they both shared the same text type, number of words, and difficulty.

Their findings suggest that MTPE is more productive than HT: it produces fewer errors than HT, and participants perceived the amount of time and work that was saved through MTPE to be greater than HT. Participants also reported a strong desire to learn PE skills. Although the findings of the study were very promising for MTPE implementation in the classroom, the study did not seek to measure the effectiveness after a PE course but rather held the experiment in one session which, therefore, provides results drawn from a single session of data collection. In addition, the study relied mainly on quantitative data and did not employ a mixed-method approach, which would have provided an in-depth explanation of the results drawn from the quantitative data. The current study seeks further explanations by linking the results from the quantitative data with qualitative findings from the focus group discussions held before the intervention, and the results from the retrospective pre-test survey that were administered to the participants after the intervention.

Daems's (2016) study, which was reviewed in the previous section about opinions, also attempted a comparison of speed between HT and MTPE in two groups of participants: MA-level translation students and professional translators who translated general texts from English to Dutch. The results suggested that MTPE was faster for both groups and that productivity gain was in favour of the groups who used MTPE. The current study shares similar profiles of the students with Daems's study in the fact that they do not have any prior experience in PE. However, Daem's study did not involve a PE course, nor did it use domain-specific texts. Its significance for the current study stems from its scope, which covered students' attitudes, translation process and product. Thus, including it is important as it is one of the few studies that correlate the variables of attitude, process, and product.

Although all of the studies reviewed so far showed increased productivity in the case of MTPE, other studies, such as Garcia (2011), have reported that this is not always the case. Garcia's participants were asked to post-edit MT output without using a dictionary for reference. Therefore, the author anticipated that hesitancy due to the inability to look things up would affect the duration of the task. The current study investigates the productivity of students using MTPE, with the students allowed to use a dictionary, in order to find out whether this increases productivity, especially as the researcher has personally observed that the students in the current study tend to over-rely on dictionary use.

This section has reviewed the literature on MTPE speed and productivity in different settings, and with different participant profiles, as well as studies comparing HT and MTPE performed by

translation students. Generally, there is a scarcity of studies on English-Arabic speed or productivity that answer the research question which asks about the differences in the process between HT and PE in an English-Arabic educational setting. It is true that research has shown that MTPE increases the productivity of translators of different language pairs, including the English-Arabic language pair. However, the literature is lacking in studies concerning the productivity of MTPE among native Arab undergraduate students in the translation classroom. The current study will try to fill this gap through evaluating the effectiveness of MTPE productivity in an undergraduate translation classroom and, if productivity gain is proved through the results, recommending guidelines for MTPE training to succeed. Also, although specifying the gender of the participants in the study was circumstantial rather than intentional, the current study provides results of benefit to gender-specific studies, as all the participants are female translation students. The next section will discuss TQA and reviews studies on translation quality, with a special focus on common errors in Arabic translation and Arabic MT.

### 2.4.5 Translation Quality

In addition to the translation process, translation product or ''quality'' has always been a core interest in PE research. But according to research focusing on TQA, the evolution and widespread embrace of translation technologies have paved the way for an abundance of differently operationalised definitions of translation quality (Gaspari et al. 2015). Researchers interested in PE training have used different approaches and criteria to assess translation quality in attempts to provide findings that may help improve the quality of PE outcome and training.
In an attempt to answer the third RQ, on the differences in the product between HT and PE, the present study evaluates the quality of post-edited Arabic translations by students after PE training; a comparison is made between an HT task and an MTPE task through applying two assessment approaches, i.e., scoring and annotation. Scoring is used to answer the first sub-question, ''Is there a difference in the overall quality between the product of HT and the product of MTPE?'', while annotation will be used to answer the second sub-question, "What are the most common errors in HT and MTPE tasks in the language pair English-Arabic?" Such findings could be valuable for stakeholders at translation institutes where Arabic translation is taught by providing them with details that help them decide whether or not to adopt PE training and guidelines and how to implement them. Also, although students may not be able to identify all errors, developers of Arabic MT may use these findings to improve the quality of Arabic MT output since types of errors in Arabic NMT might differ from those identified in previous studies.

Since this part of the study involves an evaluation of MTPE with regard to translation quality, we first review the different approaches used for translation quality assessment in order to provide

the background and justification for choosing the TQA approach used here. Next, in reviewing studies which compare HT and MTPE, this section notes gaps in literature that the current study intends to fill, and provides a reference for discussion of the results. Lastly, common errors in Arabic translation and in Arabic MT will be identified to provide the background to exploration and identification of translation students' errors in HT and MTPE.

**TQA Approaches**

Castilho et al. (2018) provide a comprehensive guide to TQA in *Translation Quality Assessment: From Principles to Practice* that lists and defines the strengths and weaknesses known about the different human and automated TQA approaches. According to the authors, there is a range of approaches to TQA in the HT context as well as different approaches to evaluate MT quality, although the authors suggest that the line between assessing HT and assessing MT is becoming more and more blurred. Since automated TQA lies outside the scope of the current study, we focused on TQA performed by humans. The chapter by Castilho et al. (2018) provides useful details about the different TQAs performed by humans, such as adequacy and fluency, readability and comprehensibility, acceptability, ranking, usability and performance. In addition, it includes a useful discussion about the profiles of evaluator(s) recruited to assess the translation quality, asking whether evaluators are amateur or professional, individuals or members of a group of evaluators. Such reference is useful for informing the current study in terms of deciding on the TQA suitable for its scope, and on the criteria employed when recruiting evaluators to assess the translations of the participants. Details about the chosen TQA and the criteria used to recruit reviewers in the current study are detailed later in the methodology chapter.

Studies comparing HT and MTPE have adopted differing TQA approaches. This section reviews some of the approaches used in studies that we intend to compare our results against. Through reviewing the TQA approaches of such studies, we highlight the reasons why we decided to use or refrain from using the TQA approach used in those studies. For example, Haji Sismat (2016), reviewed above, adopted the MeLLANGE typology as the error analysis approach of TQA. The author compared the translation quality of each task in order to observe the differences in quality. Errors were penalised according to their level of severity because, according to the author, depending on the error count alone can be overwhelming and not provide accurate results. Although the approach employed by the author was comprehensively justified for their scope, it falls outside of the scope of the current study because MeLLANGE typology was used to examine quality based on sentence length and TM fuzzy match, whereas the current study seeks to use a TQA approach that is appropriate for comparing the linguistic and technical features of both HT and MTPE tasks in an educational setting.

Yang et al. (2020), reviewed above, share similar scope with the current study as it compares HT and MTPE in an educational setting to evaluate the effectiveness of MTPE. It adopts the Multidimensional Quality Metrics (MQM) (Uszkoreit and Lommel 2013) as an objective tool for describing translation errors in both human and MT. According to the authors, MQM proposes two prominent metric types: accuracy and fluency, where accuracy errors relate to how well the target text represents the contents of the source text, and fluency errors are related to the language of the translation regardless of whether it succeeded in transferring the meaning or not. The TQA approach used in Yang et al. has been described as a comprehensive quality assessment tool that follows a hierarchy of error categories with a set of specifications for each metric. The MQM is similar to the TQA model used in the current study in the sense that they both check errors in accuracy and fluency. The reason the current study did not use the MQM was due to concern about the reliability of the scores given by the evaluators. Although MQM has the advantage of industrial use, the strength in the model used for the current study (*DipTrans Examiners' Marker Sheet*) over MQM is that it is easier for evaluators who have never used either model to comprehend and follow. The current study will compare the results of accuracy and fluency tests with the results of the reviewed study.

The scope of the current study requires a TQA approach that can provide a comprehensive evaluation of the effectiveness of MTPE in the translation classroom. The approach should not only provide reliable results in terms of error count, but investigate the different error types for the evaluation process to be considered holistic, while at the same time easy to follow by evaluators who have never used it before. Therefore, the current study adopts the approach to error counting used by He (2014), i.e., the DipTrans *Examiners' Marker Sheet*. The DipTrans *Examiners' Marker Sheet* is the assessment format of tests used in the Diploma in Translation (DipTrans), issued by the British Chartered Institute of Linguists. The author used this approach to evaluate MTPE tasks completed by a number of graduate students translating from English into Chinese. The DipTrans examination is available in any language pair, subject to sufficient candidates. In the version used for the current study (in 2019), there are three tasks: a general text and two semi-specialised texts. The assessment approach is designed to consider error types found in both text types, and it has generated reliable results since its inception (He 2014). Further details of its use in the current study can be found in the methodology chapter, Section 3.8.

**Studies Comparing Translation Quality Between HT and MTPE**

This section reviews studies that compare HT and MTPE tasks, while highlighting the gap in the literature of similar comparative research into the language pair English-Arabic in native Arabic

educational settings. A vast amount of research has compared HT and MTPE, especially to examine effort, productivity, and quality. A study by Krings (2001) was one of the first to compare effort between HT and MTPE. While different studies compared HT and MTPE for different reasons (e.g., Guerberof 2009: comparing MTPE output quality against translation memory output; Screen 2019: readability and comprehensibility of TTs produced by HT and MTPE; Čulo and Nitzke 2016: change of linguistic profile of TTs produced by HT and MTPE), many studies compared HT and MTPE to explore the differences in quality. This section reviews some of these studies and justifies why they were chosen. Out of fourteen studies that we reviewed in this regard, the studies by Green et al. (2013) and the study by Haji Sismat (2016) that were previously reviewed in the section about translation productivity are the only studies that compared HT and MTPE tasks which involved the English-Arabic language pair.

Haji Sismat's (2016) unpublished PhD thesis succeeded in presenting and classifying the errors that might appear in English-Arabic MT. However, it relies on the performance of learners of both English and Arabic, and does not provide a direct answer to the third RQ of the current study. However, the thesis provides a reference for comparison, as participants in both studies are translation students translating from English into Arabic using the two modes of translation (i.e., HT and MTPE). Haji Sismat conducted this comparison by running experiments on the application of MTPE in a translation classroom setting. In the phase of the study that involved comparing translation quality between HT and MTPE, the author assessed the translation quality of legal and journalistic texts. The texts ranged in length between 116 and 311 words, and the sentence length was used as a factor to run the quality analysis. When comparing translation quality, the author used the error approach (MeLLANGE error typology) to explore the error types that are commonly found in the TTs of translation students. However, the author used different texts for the different tasks, which might have added to the variables affecting the results, especially since the texts used in HT were of a different length and genre than those used in the MTPE task. The results revealed an improvement in the quality of TTs for those students who used MTPE. Nonetheless, syntactic and lexical errors were the most problematic in the post-edited texts. Another major finding was that students tended to overlook errors that were caused by cross-linguistic influence, such as article, gender, number and the Arabic conjunction ''*wa*'' (which means *and*). The author provides some practical guidelines such as PE strategies and PE rules to use in order to achieve publishable quality. Similarly, the current study uses an error approach to analyse results from translating technical texts, but the model used in the current study fits more with its aim (see Methodology 3.8). The current study also benefits from comparing the results from error count and error type with the results of Haji Sismat's study, because the trainees in his study generated TTs using the language combination English-Arabic, even though the participants were non-native

speakers of Arabic. This choice was made because of the lack of studies that used generated Arabic TTs.

Daems's (2016) research is quite different from the current study, especially in that it recruited students as well as professionals, used a different language pair, and used a different TQA model; however, its significance stems from the fact that it is one of the first studies to examine translation process and product through comparing HT and MTPE while exploring the attitudes of the participants. In the phase of the study that tackles translation quality, the author investigated translation problems by comparing HT to MTPE TTs. The author assessed the translation quality using a fine-grained quality assessment metric developed by Daems et al. (2013). Although the findings revealed that PE improved the task speed, they did not reveal significant differences in quality between HT and MTPE. The author compared the quality of translations produced by MA translation students, who were more strongly motivated to master the skills of translation, but who did not receive PE training. The current study hypothesises that its participants hold a similar urge to master the profession, and they already share the no-experience profile of the participants in Daems's study. Daems's study did not involve any form of PE training. The aim of the present study, however, is to find out whether BA students who receive PE training reveal different results in terms of translation quality. This study will similarly compare HT to MTPE but in a different language pair (English-Arabic) in order to find out whether MTPE produces quality comparable to HT within less time.

The following two studies receive their importance from their aims and scopes as they both examine translation productivity, quality, and the attitude of students in educational settings. They were both reviewed in the previous sections on opinion and productivity, and the sections that compare translation quality between HT and MTPE in them are reviewed here.

Yang et al. (2020) reported on the effectiveness of MTPE through measuring quality scores and quality errors. They used two technical texts for the HT and MTPE tasks based on pre-set criteria that they adopted from Daems (2016). The authors checked and confirmed that there were no Chinese translations of the chosen texts available online. None of the participants had prior experience in PE. Participants were allowed to use external resources such as e-dictionaries and they were instructed to produce the best translation possible without time pressure. The results revealed that the students scored higher in the MTPE task than in HT. The error analysis showed a higher number of accuracy errors than fluency errors in the MTPE TTs. Yang et al. provides a solid reference for correlating translation productivity, quality, and attitudes of users in an educational setting. However, it recruited one group only that served as the experiment group. The current study aims to obtain more accurate results by comparing two groups of students (of comparable

level in translation skills) where one group uses HT in the pre-test and the post-test, while the other uses HT in the pre-test and MTPE in the post-test. In addition, although both our study and that of Yang et al. (2020) use the same MT system, i.e., Google Translate, the current study enforces time pressure to mimic the work environment in the real world, whereas Yang et al. did not set a time limit for their participants.

Jia et al. (2019) conducted a study of similar scope to the current study, in which they compared translation from scratch with MTPE of general and domain-specific English texts into Chinese in a classroom setting. A total of thirty first-year MTI students at a Chinese university participated in this study. The authors used general texts and a dishwasher manual as a sample for technical texts. The adequacy and fluency criteria developed within TAUS's Dynamic Quality Evaluation Framework (TAUS 2013) was used to assess two selected translations (one general and one technical) due to time and cost constraints. Each one of the texts contained eleven sentences. However, the authors give no criteria or reasons for why or how these two texts were chosen. The authors made up for the small sample size of TTs by recruiting four evaluators. The findings revealed that the MTPE task generated TTs which were equivalent in fluency and accuracy to TTs resulting from HT, as there was no significant difference between HT and MTPE in the average total scores obtained from translating technical texts. Jia et al. concluded that '[t]here seems to be a tendency for post-editing to deliver comparable or even better translation quality in comparison to from-scratch translation' (2019: 63). Although the authors acknowledge that the sample size was too small for generalisations to be made, it would be interesting to compare the findings of their research with those drawn from the current study, because both studies run comparisons of HT and MTPE. Nevertheless, it is important to highlight that the study by Jia et al. used the same experimental procedure as Yang et al. (2020) reviewed above, as they gave different texts to different groups, and one group used HT to translate their text, while the other used MTPE to translate a different text. It is true that the authors used measures to ensure texts were comparable, but the fact that the texts, the participants, and the translation methods were all different would make it hard to decide whether statistically significant differences were due to translation method, participant skill levels, or differences in text difficulty.

The work of Garcia (2011) is important for the current study due to its similarity of procedure, since it also recruited two groups (control and experiment). Yet the difference lies in that the control group had to translate from scratch first and then use MTPE, whereas the experimental group used the MTPE method followed by HT. Garcia (2011) assessed the quality of post-edited texts produced by Google Translator Toolkit from English to Chinese (14 students) and from Chinese to English (21 students). In common with our study, the students in Garcia's research were asked to produce the best translation quality whether they were using HT or MTPE. For this

purpose, each group had to translate a 250-word passage within one hour, regardless of the method they were using. The author did not provide details about the difficulty of the texts; however, the profiles of the participants were identified as third-year BA translation students and first-year MA Chinese students who were studying conventional English translation. In contrast to the findings of Daems (2016), Garcia revealed that although productivity was not significantly improved, the quality was indeed better in the case of PE in both translation pairs. The results of the comparison showed an increase in quality in the case of MTPE, regardless of the translation direction, text difficulty, or translator's level of performance. Garcia (2011) concluded by insisting that PE training is the way forward, especially as MT systems are constantly improving. In a similar sense, the present study attempts to find out whether PE training would improve the quality of translations produced by translation students when using Google Translate in one translation direction (i.e., English-Arabic). This, therefore, answers the research question of whether MTPE is the way forward in undergraduate English-Arabic translation classrooms.

Although numerous studies have undertaken inspiring comparisons between HT and MTPE to evaluate the quality of MTPE, and other studies have examined the errors in Arabic MT, evaluating English-Arabic MTPE in an educational setting has seldom appeared in the literature. In addition, there has been little research into the effectiveness of MTPE in undergraduate translation classrooms where English-Arabic in the language pair in use. This study, therefore, endeavours to bridge this gap by comparing HT and MTPE in languages from two different families. This evaluation hopes to determine the effectiveness of MTPE training in the classroom of English-Arabic, especially at an undergraduate level.

**Common Errors in Arabic Translation and MT**

The current study seeks to identify and then compare errors in Arabic HT and MTPE and to answer the sub-question of RQ3: What are the most common errors in HT and MTPE tasks in the language pair English-Arabic? Therefore, a search for studies that might have already identified these errors was undertaken to provide a reference for comparison. The most recent study identified was ''Lexical Problems in English to Arabic Translation: A Critical Analysis of Health Documents in Australia'' by Alhihi (2016). This study attempted to identify the most common lexical errors in Arabic translation using five Arabic health documents that were translated from English. The documents had been translated by government-approved professional translators. For the analysis, accredited Arabic translators assessed the documents using a functionalist-based error analysis that focused on semantics. The findings revealed that the types of errors encountered were addition, omission, compounds, synonyms, and inconsistencies in technical terms. Alhihi's findings provide a recent reference of comparison for errors in HT found in the current study.

However, although Alhihi's study is quite recent, it refers to studies that date back to the 90s and the early 2000s (e.g., Farghal 1995; Saraireh 2001). This might be an indicator of the lack of research into translation errors in the Arabic language, and, although outside of the scope of the current study, our results could therefore add to the literature on errors in Arabic translation.

As for common errors in Arabic MT, they were an interest to a number of researchers who have evaluated the quality of Arabic and English MT output resulting from the English-Arabic-English language pairs (e.g., Abdelaal and Alazzaiwe 2020; Abu-Ayyash 2017; Condon et al. 2010). Authors have highlighted issues with errors in the Arabic MT and they have called for developments in Arabic MT systems in their recommendations. However, a significant improvement in the quality of Arabic MT has not occurred since their recommendations which suggests that introducing PE of Arabic MT output if Arabic MT is meant to be exploited professionally and commercially may be a resolution that cannot be ignored.

Therefore, to set the background for Arabic MT error analysis, a search for studies that identified such errors was undertaken. Izwaini (2006) evaluated the output of three different systems: Google, Sakhr and SYSTRAN. According to the author, the problems in English-Arabic MT translation can be classified into three categories: problems of lexis, problems of grammar and syntax, and style and spelling. In addition, the author found that deletion was the major problem found in MT from English into Arabic. In a similar sense, but using a different approach of classification, Al-Samawi (2014) conducted a study to classify the errors found in English-Arabic MT based on syntactic, grammatical, and semantic errors. The results revealed that grammatical errors have the highest percentage of occurrence, with 47.5% of the total error count, followed by sematic errors, with 37.4%, and finally, syntactic errors, with 15.1%. Such error types will provide a reference for identifying errors in MTPE for this study, since it was not possible to find any MTPE error classification in the Arabic language.

Furthermore, Al-khresheh and Almaaytah (2018) conducted a study to explore the linguistic errors faced by MT when translating English into Arabic through using proverbs as a sample for translation. They also wanted to confirm the need for the human intervention to fix the TTs, i.e., the need for PE. They used GT to translate a number of randomly selected English Proverbs into Arabic. After quantitatively analysing the TTs, the authors found that GT faced some linguistic obstacles when transferring meanings from English into Arabic. In terms of comprehension and accuracy, they found that words of multiple meanings presented countless challenges in online translation. In addition, using GT provided misinterpreted TL equivalents, literal translation, wrong word order, and inappropriate lexical words in Arabic. The present study will scrutinise these findings through running an analysis of error types found in translation students' tasks that

were carried out through MTPE and to evaluate the effectiveness of MTPE as a training medium through confirming whether it helped in eliminating those errors.

Alanazi (2019) identified the challenges in Arabic MT output as morphological, phonetic, and phonological. These challenges were listed as the reasons behind the negative perceptions of MT output among the translators of Arabic. The author used an experiment in which translators post-edited both Arabic TM suggestions and Arabic MT output. The results suggested a tendency within Arabic language translators to edit TM and post-edit MT suggestions. Although the translation students in the current study have no previous experience in PE, like the participants in Alanazi's study, the current study attempts to find out whether the quality of the final product is improved after the students take a PE course which addresses the expected challenges in Arabic MT output.

More recently, Al Mahasees (2020) conducted a study to evaluate the Arabic output of three MT systems (Google Translate, Microsoft Translator, and Sakhr) using holistic and error analysis of the entire translation as well as collocations as a special case. The analysis employed TAUS adequacy and fluency scales for the holistic analysis, and criteria composed of orthography, lexis, grammar, and semantics at a text level for the error analysis. The analysis was run at the beginning of 2016 and at the end of 2017, in terms of adequacy and fluency, the study found that Google Translate was the best performing system, followed by Microsoft and Sakhr. Although Google Translate showed the best improvements between the two-time intervals, the study revealed that all three systems have produced errors in capitalisation, punctuation, numbers of months, omissions, additions, mistranslation, agreement, wrong word choice and word order. The author concluded by suggesting that although Neural Machine Translation (NMT) is expected to provide better results when compared to hybrid systems, post-editing is still required for the final translation to be acceptable for the end user. The current study attempts to identify the errors that still exist after the human error-fixing, i.e., after post-editing, and to determine whether PE improves the quality of the translated texts resulting from an NMT system (Google Translate).

## 2.5 Summary and Concluding Remarks

This chapter has presented an overview of several aspects related to MTPE training (opinions, productivity, and quality), and provided relevant information about the literature available on each of them. After locating 'translator training' within 'translation studies', a summarised chronology of translator training was given, highlighting pedagogical change in the field, i.e., the adoption of MT training in order to keep up with the growing industry, while noting that no active measures have been taken in the Arab world. A summarised overview of translation pedagogy

was provided also, focusing on the globalisation that has motivated the interest in MT within translation pedagogy, and the differences between 'translation training' and 'translation education'. The conceptual framework reviewed several evaluation models of learning, and key concepts were evaluated in the current research in order to arrive at a holistic evaluation of MTPE training. In addition, some of the key studies on opinions about MTPE, and studies that compared productivity and quality in HT and MTPE were reviewed.

As mentioned in the research aim, as well as in the recommendations of some studies in the literature review, MT training and PE have been repeatedly encouraged in translation training programmes that teach Arabic translation, but without further evaluation or development of this training. This has helped to identify research questions that were deemed worthy of exploration. The evaluation model of learning that has been chosen here (the Kirkpatrick model of learning evaluation) synchronises with the recommendations in the literature in the fact that it combines the evaluation of opinions with that of productivity and quality in order to arrive at a comprehensive evaluation of the effectiveness of the training programme.

The next chapter will describe the methodological approach adopted in order to seek answers to the research questions of the Methodology

# Chapter 3 Methodology

## 3.1 Introduction

This chapter is devoted to the research design to study the three aspects of this research (exploring students' opinions, comparing translation productivity, and comparing translation quality). Therefore, this chapter provides an overview of the methodological approach and methods selected in this thesis. Section 3.2 provides a research summary that includes RQs, data sources, instruments, and methods of data analysis. Section 3.3 tackles the philosophical assumptions underlying this thesis. Section 3.4 describes the experimental design that was applied in this study followed by an introduction to the features of mixed-methods design and a rationale for its selection while section 3.5 describes the research settings. Section 3.6 details the criteria used for recruiting students in this study. The training programme and TQA are covered in section 3.7. Section 3.8 illustrates the TQA employed in the current study. Section 3.9 describes data collection methods, while data analysis is detailed in section 3.10. Lastly, before ending the chapter with a reflection on the researcher position in Section 3.12, Section 3.11 presents a brief note on ethical considerations.

## 3.2 Research Summary

In Table 3-1 below, a summary of the research is provided where all the RQs, justifications for each RQ, the gathered data for each RQ, research instruments, and method of data analysis are provided.

Table 3-1 Research summary

| | Research Question | Data Source | Instrument | Data Analysis |
|---|---|---|---|---|
| **Opinions** | **RQ1:** What are the differences in students' opinions about HT and MTPE?<br>1.1. How rewarding is MTPE compared to HT?<br>1.2. How useful is MT output according to translation students?<br>1.3. Which translation method is perceived as being faster?<br>1.4. How is the quality of both methods of translation perceived?<br>1.5. Which translation method is the most preferred? | Students' feedback (transcribed recorded data) | Focus Group Discussions | Thematic analysis |
| **Opinions** | All previous sub-questions +<br>1.6. Is there a difference in students' opinions before and after the intervention? | Students' feedback (survey responses) | Retrospective Pre-test Survey | paired-samples *t*-test |
| **Productivity** | **RQ2:** What are the differences in the process between HT and MTPE?<br>2.1. Is MTPE faster than HT? | Total duration in minutes | Translation task | paired-samples/independent-sample *t*-tests/ANOVA |
| **Quality** | **RQ3:** What are the differences in the product between HT and MTPE?<br>3.1. What are the most common errors in HT and MTPE tasks in the language pair English-Arabic? | Evaluators' annotations | Translation task | Total count of error types |
| **Quality** | 3.2. Is there a difference in the overall quality between the product of HT and the product of MTPE? | Total score out of 100 | Translation task | paired-samples/independent-sample *t*-tests/ANOVA |

## 3.3 Philosophical Assumptions

Saldanha and O'Brien (2014) suggest that the success of the methodology in addressing the research questions depends in the first place on how well the methods suit the research questions and the overall aim of the study. Behind every study, there is a research philosophy to conduct the study in an appropriate and effective manner that assumes a specific worldview. It is useful in the sense that it helps the researcher choose the appropriate strategy and data collection methods (Saunders et al. 2015). According to Saunders et al.:

> The term research philosophy refers to a system of beliefs and assumptions about the development of knowledge. Although this sounds rather profound, it is precisely what you are doing when embarking on research: developing knowledge in a particular field. The knowledge development you are embarking upon may not be as dramatic as a new theory of human motivation, but even answering a specific problem in a particular organisation you are, nonetheless, developing new knowledge. (2015: 124)

Therefore, in this section, I am describing the philosophical stance in the current research. My starting point was my intention to fulfil the aim of the research as well as answering the main RQ (How effective is MTPE training in a female undergraduate translation programme in Saudi Arabia?) rather than delving into the philosophical issues associated with every research approach. According to Hughes and Hayhoe (2009), the use of quantitative methods, or qualitative methods alone may not be sufficient in answering the research questions in a study that may, otherwise, be answered through using both methods. When attempting to answer the research questions of the present study, I needed to gather qualitative data such as focus group discussions to explore the opinions of students and to inform the design of the teaching intervention that would be offered to the students. In addition, I needed to gather quantitative data such as measuring the translation productivity, TQA, and a retrospective pre-test survey which are required for the analysis after the final phase of the study is concluded. Thus, as I was learning more about the variety of the philosophical worldviews, pragmatism seemed the most appropriate world view for tackling my study. In pragmatic research, the research question is the most important factor to determines the research philosophy where both qualitative and quantitative methods can be combined within the scope of a single research as long as they are attempting to answer the research question (Saunders et al. 2015).

Undertaking this pragmatic approach to research has the following three strengths. Firstly, its particular usefulness in evaluation and survey (Patton 2002) because it has broader focus than the single-method design, thus it gathers more data in different modes about the phenomenon. Secondly, it can provide insight into the complexity of a social phenomenon by producing findings that illustrate that complexity (Giddings and Grant 2006). Thirdly, the breadth of its findings that may bring value to the research process itself by highlighting the shortcomings in each of the methods that were used, and ways used to compensate for them (Giddings and Grant 2006). Nonetheless, despite the strengths associated with this approach, the major challenge that I faced when adapting this philosophy, is that it usually takes experienced researchers to be able to handle this approach. Therefore, I had to consider the strengths and the weaknesses, and I concluded that the strengths of undertaking this approach of research might prove superior once I attempted to tackle its limitations.

I tackled each of the challenges of this research approach by the following: (1) It consumes more time, both at the beginning (planning and negotiation) (Giddings and Grant 2006), this was resolved by the thorough consideration of different research approach options and discussing them with the research supervisors (see appendix H- Visited Research Designs). (2) The data analysis of this type of approach also takes more time; this issue was solved through utilizing technology; by using computer-assisted qualitative data analysis software (CAQDAS) such as Excel for the analysis of the qualitative data, and SPSS for the quantitative data, it was hoped that time saved during the data analysis would allow more time for the later interpretation and discussion of findings and results.

In addition, based on the aim of this study, the logical research framework has to be of an evaluative nature. Thus, for the purpose of organizing the different phases and processes of this study, and in order to explore the relationships between the different levels of the evaluation model used, I followed the structure outlined by the Kirkpatrick's Model for Learning Evaluation (See section **Error! Reference source not found.**) mainly to organize the logical flow of data gathering and analysis about different levels of the study: opinions, process (translation productivity), and product (translation quality). An additional benefit of utilizing the Kirkpatrick Model for Learning Evaluation is that it can be applied in different pedagogical settings, and its design follows a logical chain so that if the training programme does not meet its required goal, it is easier for the researcher to find out what went wrong and in which level of the evaluation process (Kirkpatrick and Kirkpatrick 2016).

The pragmatic nature of this research enabled me as a researcher to mix different components of a single design that is expected to provide the most appropriate way to address each research

question while the evaluation model provided the logical as well as the analytical structures of the research. Advocates of pragmatism highlight that the major advantage of the mixed-methods approach is that it emphasises the strengths of each method, while the weaknesses are reduced with proper planning (Creswell 2003; Teddlie and Tashakkori 2006).

## 3.4 Research Design

Yin defines research design as 'the logical sequence that connects the empirical data to a study's initial research questions and, ultimately, to its conclusions' (2017: 20). Yin also uses the phrase 'the logic of the design' (2003: 14) to refer to the process in which research questions are linked to data as a part of a comprehensive research strategy. As I applied a pragmatic approach to understanding and answering the research questions, a sequential exploratory mixed-methods design, embedded within an experimental design resulted. The following subsections provide further details on the applications of these two designs.

### 3.4.1 Experimental Design

To evaluate the effectiveness of MTPE training in an undergraduate translation program, an experimental design was selected. Based on pragmatism through which the researcher seeks to identify a solution for a real-world problem, the experimental design assists in identifying the effect of the MTPE training on the variables of interest (i.e., students' opinions, translation productivity, and translation quality). Therefore, it was expected that the MTPE training would show some sort of an impact on all or some of the three aspects that are being researched. The experimental design allows the influence of the MTPE training on translation students to be determined. Saldanha and O'Brien (2014) state that when a researcher seeks to determine cause and effect, an experimental design is an ideal methodology for the investigation 'if X happens, then what is the effect on Y?' (2014: 15). In addition, Randolph (2008) notes that the experimental design can be used when a researcher seeks to compare the results or findings that emerge after altering a phenomenon and monitoring its impact. In experimental design, the researcher intentionally manipulates the independent variable(s) to determine the effect of these changes on the dependent variable(s). In the present study, the independent variable is the MTPE training, while the dependent variables are the translation students' opinions, the translation productivity, and the translation quality.

The nature of the evaluation intended in this study required assessing the effectiveness of MTPE training on the opinions, productivity, and quality of undergraduate female translation students, and one way is to expose the students to MTPE training in the classroom. To determine the

effectiveness of the teaching intervention, I compared the presence and absence of a MTPE training course on the three different variables within the present study. Therefore, the experimental design was the most appropriate methodology to investigate and answer the research questions, as it could provide explanations regarding whether the presence of MTPE training could affect students' opinions, their productivity, and the quality of their TTs in the process of learning to translate.

The pre-test-post-test control group and experimental group design with repeated measurement was used in the current research study. In this design, translation students were randomly assigned to either the control or the experimental group based on cluster sampling, i.e., the sampling unit is the group or cluster of participants. The main reason for the use of the cluster sampling technique was to evaluate the effectiveness of the intervention on the natural structure of students that make up a classroom (Saldanha and O'Brien 2014). Participating students came from two different cohorts that were studying the same level in the translation program. In the recruitment process, the first cohort that was recruited was randomly assigned as the control group, and the second cohort was assigned as the experimental group. The average number of students in a cohort in the translation programme is 35, of whom 29 students from 'cohort A' consented to participate in the study (later assigned as Control Group, or Ga), and 31 students from 'cohort B' consented to participate in the study (later assigned as the experimental group, or Gb). Regarding the random allocation process, all translation students who agreed to participate in the research were assigned a number on an Excel spreadsheet. The following formula (adapted from He 2014) was used to generate random IDs for students (Group ID + N.+ student serial number as it appeared on the excel sheet). Next, students assigned to the experimental group received the intervention (MTPE training course), while the students in the control group received conventional translation lessons (HT). The following formula explains the pre-test- post-test students IDs in both groups:

**Control group:** Ga

An example of a student ID in the control group is: GaN12 (student number 12 in the control group).

**Experimental group:** Gb

An example of a student ID in the experimental group is: GbN27 (student number 27 in the treatment group).

**Pre-test:** Ta

Example: The pre-test result of student number 5 in the control group would be (TaGaN5)

**Post-test:** Tb

Example: The post-test result of student number 9 in the experimental group would be (TbGbN9).

The selection of the experimental design is supported by two factors: The first is that this design allows for comparisons between two situations (i.e., the presence and absence of MTPE as a method of translation). Therefore, two groups are required to identify the effectiveness. The second factor is related to the strength of the design, which takes prior performance of both the control and the experimental groups into account and helps to determine whether there are any differences before and after the intervention is introduced (Saldanha and O'Brien 2014). This is useful in determining the level of translation performance (both process and product) in addition to exploring the opinions for both groups before introducing the MTPE training. The identification of prior performance for both groups hopes to reduce the impact of other variables such as the effect of the data collection environment (language labs), or the use of keyboards and monitors which may affect the routine behaviour of students. Such identification of prior performance may increase the confidence in the findings and ensure that if any differences are found that they are due to the independent variable (Saldanha and O'Brien 2014).

I was aware that controlling all variables in this research was difficult. Saldanha and O'Brien (2014) point out that although the experimental design is to be carried out in a controlled environment, this is not always practical in humanities and social science research. They suggest that '[c]omparable groups of translators or translations may simply not exist' (2014: 15). Therefore, despite the fact that I have taken all possible measures to control the different variables between the control and the experimental groups, ensuring a fully controlled experimental setting was not achievable due to the natural distribution of the participants (the cluster sample) and the fact that the levels of their educational and experience levels might not be totally even. As previously mentioned, there were several uncontrollable variables in the present study, such as students' experience using computers and content-editing. Due to the limited period allocated for data collection, I considered providing students with tutorials without the ability to ensure that they actually practiced the skills in such tutorials (which in itself is an uncontrollable variable). Therefore, some tutorials on the skills required for the study were provided to students in both groups by emailing them some videos and guidelines (e.g., software skills, useful keyboard shortcuts, and basics of touch-typing) ensuring that they knew how to deal with computers (in both groups) and with MT output when post-editing (in the experimental group).

### 3.4.2 Mixed-method Design

As far as the nature of this study is concerned, it is primarily evaluative, seeking to answer the question: how effective is the program? (Duff 2018). The evaluation itself is of a relational nature, as it goes beyond pure description to find the causal patterns among the different variables (what is the effect, if any, of the independent variable, MTPE training, on the dependent variables, translation productivity and translation quality, and are students' opinions changed after MTPE training?) in an effort to yield explanations about the effectiveness of translator training through the use of MTPE. By conducting this research following these measures, I intend to explore the themes, models or principles that may emerge from this exploratory study.

The general aim of evaluating the productivity of translation is to study the extent to which the MTPE training achieves its objectives (i.e., allowing the participant to achieve a faster outcome than HT). The aim of running a translation quality assessment is to study the value of the MTPE training as reflected by the scores of the translated texts (TTs). The aim of exploring students' opinions receives its importance from previous results in the literature which suggest that even if post-editing is found to be faster without compromising the quality, it is still important for translators to show positive attitudes towards technology for it to be deemed useful (Alotaibi 2014; Daems 2016). Combined, these three aims can help decide on the MTPE training program's effectiveness in the first place. Then modifications, improvement and programme certification and adoption may be generated from the results of this study. Such evaluation requires a design that is flexible and integrated because the learning experience is originally composed of all three aspects that constitute the objectives of the current research study, hence I used a mixed-method design. Yet, evaluating the different aspects of the learning experience separately provides a clearer vision of the resulting themes and principles, and Kirkpatrick model of learning evaluation provides the smooth flow of the evaluation process.

A strength of the mixed-method design when exploring students' opinions in the present study is that it allowed me to triangulate the data sources (focus group discussions and a survey) which sought convergence throughout qualitative and quantitative methods (Saldanha and O'Brien 2014). According to Nastasi et al. (2007), evaluation research should involve the triangulation of qualitative and quantitative methods in order to examine acceptability, integrity, and effectiveness of intervention methods as both a formative and summative process. For these reasons, adapting mixed-method design in this study was useful due to the complexity of evaluating the effectiveness of MTPE training in the undergraduate translation programme targeted in this study.

Another strength of the mixed-method design is that it reduces the weaknesses that arise from using a single method. According to Creswell, 'recognising that all methods have limitations, researchers felt that biases inherent in any single method could neutralise or cancel the biases of other methods' (2014: 15). Thus, the combined design of this research study provides a comprehensive evaluation of the research enquiry. In addition, the use of different methods corroborated the findings in each of the different stages of the research. For all these reasons, different data collection methods were used (focus group discussions, the calculation of translation total time and scores, identifying translation error types, and the use of a survey).

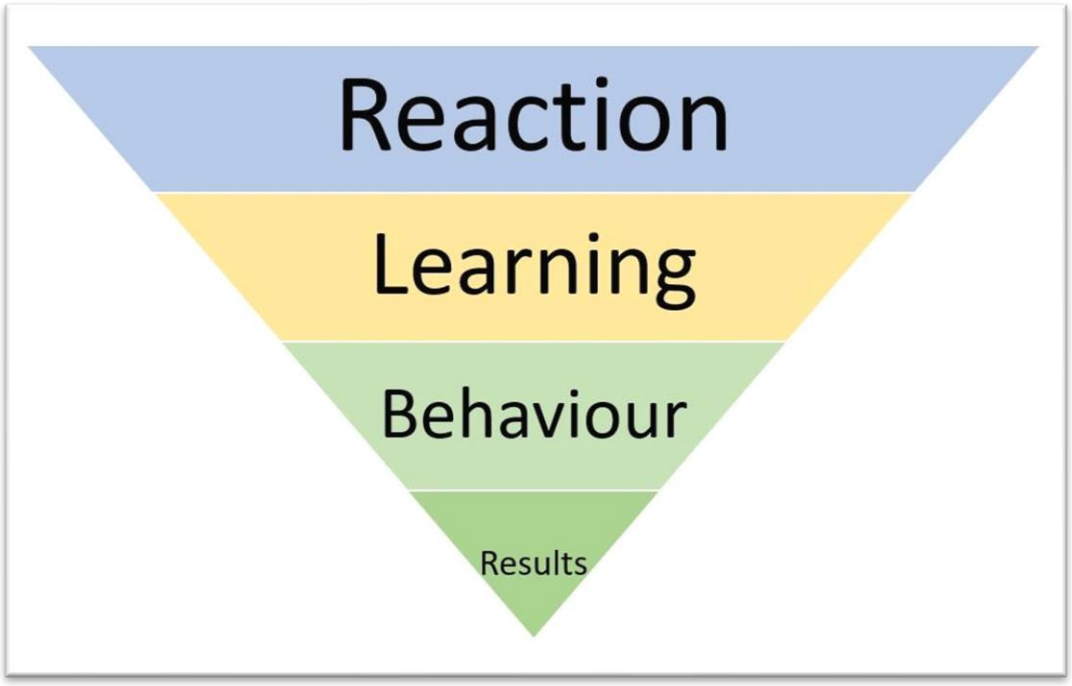
Nonetheless, despite the several benefits that the mixed-method design has, there are a number of weaknesses that every researcher has to consider when using this approach. According to Creswell and Clark (2017), the researcher has to have a deep understanding of the different data collection methods as otherwise, the validity and reliability of the research would be jeopardized. I attempted to overcome this issue through attending a number of online tutorials and reading a number of PhD theses and how other researchers ensured the validity and reliability of their research were maintained. In addition, I piloted some of the instruments and conducted several reliability and validity steps in the research. A further weakness in the mixed methods approach is that the analysis and interpretation of results would need a considerable amount of time and effort in order for the researcher to focus on the research study. Early planning and attending tutorials of online tools (e.g., SPSS for statistics, and Excel for thematic analysis) allowed me to focus on overcoming this weakness.

Data is also analysed sequentially or concurrently. In this study, data was analysed sequentially (qualitative then quantitative), however, results weight was more reliant on the quantitative analysis because it provided results to

When using a mixed-method approach, there are two ways to conduct the research: parallel or sequential (Creswell 2014). In the parallel design, quantitative and qualitative data are gathered simultaneously, and the findings are reported separately and may not necessarily relate to nor confirm each other (Teddlie and Tashakkori 2006). In the sequential design, however, the researcher starts the data collection by conducting the qualitative phase followed by the quantitative, or vice versa (Teddlie and Tashakkori 2006, Cameron 2009). The results of the first phase usually inform and help with the planning of the second phase of data collection (Creswell 2014). Because of the evaluative and exploratory nature of the current research study, and when applying the order of the levels of Kirkpatrick model of learning evaluation, the sequential design seemed more appropriate to fulfil the purpose of the current study, as supported by researchers in the area of learning evaluation (e.g., Duff 2018). According to the details of Kirkpatrick's model

of learning evaluation (Kirkpatrick and Kirkpatrick 2016), a model of learning evaluation should involve an assessment of the various levels that are involved in the training program; in the current research, these are: reaction (students' opinions), learning (the MTPE course), behaviour (the process: translation productivity), and finally, results (their quality scores). The Kirkpatrick model of learning evaluation (See Figure 3-1 below) highlights the importance of examining each of the levels separately in order to decide whether each level was effective (Kirkpatrick and Kirkpatrick 2016).

Figure 3-1 Kirkpatrick's Model of Learning Evaluation

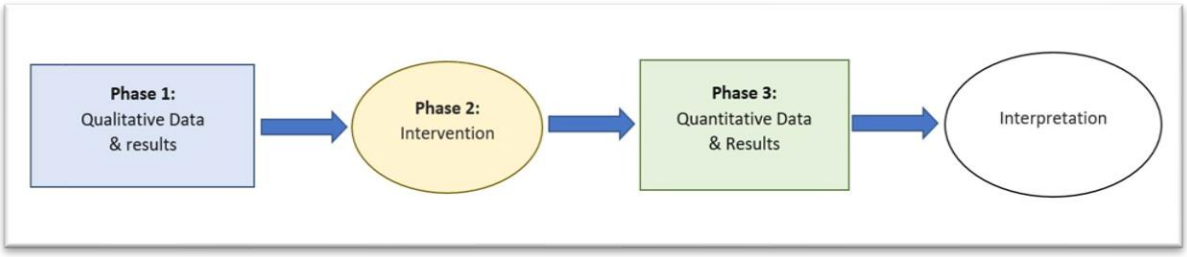



Teddlie and Tashakkori (2006) state that there are two commonly used sequential designs used for mixed methods: exploratory and explanatory. The exploratory design begins with qualitative data collection, followed by quantitative data collection, and is typically used when the researcher seeks to explore the phenomena and builds on the findings by creating and testing new instruments. On the other hand, the explanatory design is used when quantitative data is gathered in the first phase followed by qualitative data, which are used to explain the relationships and examine the results, in the second phase (Teddlie and Tashakkori 2006). For the purpose of answering its research questions, the current study utilised the strategies of an exploratory sequential mixed-method design (with multiple units of analysis) that was proposed in Teddlie and Tashakkori's (2006) typology of mixed methods research design (see Figure 3-2 below ) The exploratory study approach included using a variety of research strategies such as:

focus groups discussions (RQ1, exploring opinions), a translation task (RQ2, measuring productivity and RQ3, translation quality assessment), and a survey (RQ1, triangulating opinions).

The proposed design is composed of three phases (phases are numbered chronologically, and color-coded in a way that represents the different phases of the Kirkpatrick model of learning evaluation (Figure 3-2 below): In Phase 1 (qualitative), students took part in focus group discussions where the acquired data were qualitatively analysed, part of the data gathered in Phase 1 (i.e., students training needs) were used to design the contents of the MTPE course which was taught in Phase 2 of the study. A pre-test (quantitative) was given to the students right before Phase 2 began. Then in Phase 3 (quantitative), the post-test to investigate the differences in process (translation productivity) and product (translation quality) between the experimental group (students who took the MTPE training course) and the control group (students who studied the same translation course but in the conventional HT method) was completed by the students. In Phase 3, also, a retrospective pre-test survey was administered to students in the experimental group immediately after finishing the post-test.

Figure 3-2 Sequential mixed methods (Teddlie and Tashakkori 2006)




In more details, a mixed approach was selected in Phases 1 and 3 to investigate translation students' opinions towards MTPE. Firstly, focus group discussions (FGDs) were adapted as a method to collect as much qualitative data from translation students before the intervention. Secondly, a post-intervention retrospective survey was used to triangulate and validate the results from the FGDs through quantitative data, and to explore the change (if any) in students' acceptance of MTPE.

In Phases 2 and 3, the study applied a quantitative pre-test -post-test control (Ga) and experimental (Gb) group design (Cohen et al. 2013). The main reason for using this is that it permits comparison between two different situations (the absence of MTPE in Ga and its presence in Gb), which allows any influence of MTPE on translation productivity and quality in Gb to be determined. Students in both groups took a HT pre-test (Ta) to obtain a reference value for each group. In Gb, the translation students were participating in MTPE tasks throughout the 4-

week intervention period, and they applied this method of translation in the post-test (Tb). The students in Ga did not take the intervention but studied the usual translation lessons and participated in both the pre-test (Ta) and the post-test (Tb) only through using the conventional HT method.

Furthermore, when designing the part of the study that answered RQ 3, deciding on the aspects which would assess the effectiveness was essential for the evaluation process to produce comparable results. According to Yang et al., 'Avoiding errors and reducing the risk of unacceptable consequences are essential for task accuracy and completeness. Thus, error count and error type are key assessment indicators for effectiveness' (2020: 12). Therefore, providing definitions of error count and error type and how they would be collected in this study is a must (see 3.8 below).

Finally, the choice of conducting an exploratory study is mainly informed by the research questions. Researchers carry out exploratory studies when the topic requires in depth understanding which has not been done before (Stebbins 2001). The goal of such design is to explore the problem and around it and not actually derive a conclusion from it (Stebbins 2001).

## 3.5 Research Setting

The study was conducted at King Saud University (KSU), Saudi Arabia. KSU is the oldest university in Saudi Arabia, and it was established in Riyadh in 1957. It has an academic reputation in different fields such as Islamic, Arabic Studies, Science and Languages. Since its inception, the women section at the College of languages and Translation (COLT) at KSU has been committed to preparing both translators and interpreters to fulfil the need in the translation market. The women section at COLT consists of two departments: the department of English language and translation, and the department of French language and translation. This study was conducted on female translation students enrolled on a 5-year English translation programme during the Spring and the fall semesters of 2019. The (Technical and Scientific Translation) module was selected for this study, which is originally offered as a mandatory course for all translation students in the translation programme at the university and it must be taken before graduating the program. To avoid ethical issues arising from offering the teaching intervention in a mandatory course, I offered an optional version of the course based on alternative times that were agreed upon by all recruited students. To do so, I have offered a poll on the best suitable times to meet the students for the optional course. This module focuses on the translation of texts of technical and scientific nature, and it includes several topics concerning how a translator is to tackle these text types. The mandatory *Technical and Scientific Translation* module typically consists of weekly two-hour

lectures for a period of 15 weeks. However, the optional *Technical and Scientific Translation* module that I offered was separate module that consisted of weekly two-hour meetings in the language lab for a period of four weeks.

### 3.5.1 Designing the Intervention Course and Translation Tasks

In the present study, face-to-face (f2f) instruction served as the method of teaching. It was chosen for two reasons: Because it is currently used by COLT instructors, so the students were also familiar with this method of teaching, and because teaching is likely to continue f2f. If this intervention produced practical applications for the future, it is best to base it on the normal method of delivery of teaching. The course was offered in the language lab (see Figure 3-3 below) which provided both the traditional and interactive whiteboards for me to explain certain points while teaching as well as computer stations for the students to use while practicing the assignments and taking the pre-test and the post-test. F2f instruction has also provided an environment for discussions as students shared ideas and asked questions during the period of the intervention.

Figure 3-3 The language lab where data were collected


#### 3.5.1.1 Text Selection

Technical texts are considered the main target of MT systems and they were used in many studies of MTPE that included different numbers of language pairs (Krings 2001; O'Brien 2006; Haji Sismat

2016; Yang et al. 2020). According to Kingscott (2002), a large percentage of the world's total translation is technical texts. In addition, there is an established literature on comparison between HT and MTPE that involves technical texts. Therefore, I chose technical translation as the source for the texts used in this study in order to compare results with previous research. In terms of word count and the orientation of the tasks in the e-briefs, the criteria I used in line with guidelines suggested by Koponen (2016) which state that: too problematic and not problematic texts are not suitable for MTPE and that researchers studying MTPE tasks should avoid long and complex sentences, short and ambiguous sentences, long noun phrases, and prepositional phrases which can all be problematic for MT systems. Koponen (2016) has also recommended that texts with too many or too few complex compounds or idiomatic expressions should be also avoided. Based on these criteria, I chose a total of six texts: one text for Ta, one text for Tb (see 3.5.1.2 below), in addition to four texts that were used as HT/MTPE tasks during the training period. The texts were excerpts from WikiHow and Ariel (UK website) and a collection of user guides. To my knowledge, there was no Arabic version of any of the texts at the time of conducting the study.

All English texts were meticulously chosen in the sense that they reflect some of the main characteristics of scientific and technological texts, as required by the objectives of the course. In particular, the research study focused on the translation of expository texts (Hatim and Mason 2014). This particular text type was chosen for its capability of being post-edited, and that relatively short texts of this type can be easily found. Also, this is a text type that has a straightforward style which can be easily identified and dealt with by translation students.

Several measures were taken to make sure that students could not find the Arabic translations of the articles on the web. First, the names of the text sources were not provided in the instructions of the translation tasks. In addition, on the test days, students were instructed not to search any sections of the texts online, and they were invigilated during the tests to make sure that they were not accessing any website that did not serve the purposes of the study.

#### 3.5.1.2 Test Material

The rationale behind choosing the texts for the two tests (Ta and Tb) is to mimic the format of the translation tests used for the qualification of passing the designated course based on criteria set out by COLT while observing the criteria that were previously set by Koponen (2016). These tests are usually introduced at three intervals: Mid-term 1 is administered at the end of week 5-6 (20 points), mid-term 2 is administered at the end of week 11-12 (20 points), and the final exam (60 points) is given the end of the course as a benchmark of professional standards in the translation in that specific field (in this case, *Scientific and Technological translation*). According to the criteria

of the COLT, each HT-mode examination must include an English text to be translated into Arabic of an average count of 250 words, and the allocated time for finishing the translation task is one hour.

In an effort to get varying results (if any) that can be justified by the different method of translation that was used, I followed the same criteria for material selection and used the same Ta text and Tb text in both Ga and Gb. In both Ta and Tb as well as the extra-curricular assignments, students were given technical texts that were around than 250 words each, and each group was asked to finish the translation tasks using the translation method allocated for them. Similar to in-lab assignments, test materials used in Ta and Tb were also expository in nature (i.e., how-to texts). Their purposes were either to inform, to explain a procedure, or to report an experiment (Pontus et al. 2017).

For Ta, I decided to use a text derived from WikiHow website *How to Clean the Showerhead with Vinegar* (see Appendix I- Ta) that has no Arabic version. For Tb, I used an instruction text from www.ariel.co.uk *Using a Front Load Washer* (see Appendix I- Tb). The text for Gb in Tb (MTPE) was an Arabic MT output produced by Google Translate on the day the students sat for the translation task. The MT system, Google Translate, was employed for the study because it is NMT, it is free, and because of its wide use by the students.

#### 3.5.1.3 Metrics for Text Difficulty, Length, and Readability

The Flesch Reading Ease Formula (Flesch 1948), which is a widely used tool (e.g., Yang et al. 2020; Daems 2016) was adopted to measure the texts' readability and complexity. When using the Flesch Reading Ease Formula, text difficulty is evaluated based on average sentence length and average number of syllables per word. Texts are rated on a 100-point scale. Lower scores indicate more difficult texts, and higher scores indicate easier texts. Although I made every effort to find texts that were exactly similar to maintain the validity of the results, it was not possible due to the fact that I followed criteria that were composed of five aspects: (1) texts have to be eligible for producing comprehensible MT output, (2) maintain COLT standard number of words in the text, (2) avoid too easy/too problematic sentences, (3) avoid long sentences, (4) avoid complex sentences. Therefore, after running the readability test on www.webfx.com (used by Daems 2016), I decided to use the slightly easier text in Ta because it is expected that after four weeks of training, students in both cohorts would gain more translation experience regardless of the method used. Overall, scores below suggest that the texts were of comparable difficulty and appropriate for native readers between 13-14 years (Ta) and 14-15 years of age (Tb).

Below are the major readability metrics used when selecting the texts for Ta and Tb.

Table 3-2 Comparability of texts

| Criteria | Pre-test Text (Ta) | Post-test Text (Tb) |
|---|---|---|
| Word Count | 247 | 249 |
| Flesch reading ease score | 70 | 66 |
| Readers' Age | 13-14 | 14-15 |

## 3.6 Research Sample: Translation Students

Saldanha and O'Brien point out that '[t]o make sure that the selected sample enables the researcher to answer the research question, consideration must be given to the sampling technique' (2014: 33). The sample for this study was selected from an undergraduate female translation students' population. A research population is defined as a large group of individuals, items or units that compose a sample (Gravetter and Forzano 2018). The population in this study was composed of translation students at COLT. COLT data on its student population indicate that there were 1936 translation students in the college specializing in either English-Arabic or French-Arabic translation (كلية اللغات والترجمة 2019). The sample was selected from the population of female translation students at COLT because of the complexity of gathering data from individuals of a large group coming from separated departments as males and females study in separated campuses in Saudi Arabia.

Although random sampling where each person in the targeted population has an equal chance of being selected to participate in the study is considered the 'gold standard' (Saldanha and O'Brien 2014: 3), it was not feasible in the current research for two major reasons: (1) Because the targeted population in the current study is not large enough, and (2) because the chosen sampling technique for the current research was thought to provide findings and explanations that represent the true population of translation students. The sampling technique used for this study is cluster sampling. Cluster sampling is 'the term used when the natural sampling unit is a group or cluster of individual units' (Saldanha and O'Brien 2014: 34). The reasons behind using cluster sampling are: (1) In the current study, the group is the unit of analysis for when evaluating the effectiveness, and also (2) because cluster sampling is more convenient and practical since students will be in their allocated cohort throughout the training intervention. However, in response to ethical issues, students in both natural cohorts who decided to opt out were excluded from the study.

The usual number of students per cohort in the college where the study took place is usually 35 students. Each cohort represents different educational backgrounds and expertise of students, that is, the college ensures that each cohort has comparable numbers of poorly, moderately, well, and exceptionally performing students. In an attempt to eliminate ethical issues arising from the cluster sampling after the 2 groups were assigned, students who gave consent to participate in the study were assigned numbers rather than being identified by their names (only the researcher had their names associated to their numbers), and individuals who considered opting out at any point, were allowed to do so at any point of the study. Then, the experimental group (Gb) were trained to meet the objectives of the curriculum of the translation course *Scientific and Technological Translation* through MTPE using Google Translate and Word Processor.

The complete sample began with 70 female translation students originally enrolled in the *Technical and Scientific Translation* module under the instruction of the same teacher. The sample was extracted from the student registry of the College of Languages and Translation (the number of participants was reduced to 60 students upon data cleansing (see 3.10.2.4). Saldanha and O'Brien (2014) list a number of reasons that complicate achieving sufficiently large data to produce significant results. These reasons include the lack of required types of texts, comparable contexts, or equally comparable participants. While acknowledging that the purpose of the current study is not to claim generalizability of findings, I had to recruit a sample of the translation population while considering the previously mentioned difficulty. Therefore, I learnt that recruiting a sample size of 30 is considered suitable and enough for statistical analysis in studies of an experimental nature (Borg et al. 2009). However, the reliability of the results increases if the sample size is larger than 30. Based on the previous criteria, the sample size of 60 in the current study is considered reasonable to obtain accurate results. The cluster sampling of students in cohorts A and B resulted in 29 students in Ga and 31 students in Gb.

The current study focused completely on female translation students who were enrolled in the *Technical and Scientific Translation* module. Because of the segregated nature of instruction in Saudi Arabia, it was easier for me to access female populations being a woman myself. The next section describes the selection criteria in more details.

### 3.6.1 Detailed Selection Criteria

This section elaborates on participants' knowledge of MT both prior to and after the teaching intervention. It also details the profiles of participants taking part in every part of the study.

#### 3.6.1.1 Students' Knowledge of MT

Prior to taking the teaching intervention, the participants in this study used MT systems without experience. Al-Jarf (2017) described the translation program at the College of languages and Translation (where the current study is conducted) as follows: The program offers 18 translation courses, 6 interpreting courses, and 2 courses on Computer Applications in Translation (CAT). In the first CAT course, students are taught about the components of computers, hardware, Microsoft Word, PowerPoint and Excel. The advanced CAT course introduced students to general theoretical knowledge about MT, but students reported that they have never been given the chance to use MT systems such as Systran (Al-Jarf 2017). Al-Jarf results showed that there is a lack in a range of MT-related skills, and accordingly the author recommended that students are introduced to touch typing, advanced word processing, document formatting, and the use of Google language tools to improve translation quality amongst a range of skills that current CAT courses do not teach to students.

The teaching intervention was planned and designed before data collection. The contents of the module were adjusted based on the results of the FGDs. For instance, general knowledge of MT was limited as students showed familiarity with how MT systems function and the range of available MT systems.

In the intervention, students were introduced to MT-related skills that they were expected to be familiar with and practice in Tb (see section 3.7 for more details about the teaching intervention). Although the module focused on PE as the main skill, it was inevitable to include sections that tackle MT output such as common MT error patterns in the Arabic language, stylistic features of the expository text type, evaluating MT errors, punctuation, formatting, and quality control.

The next section elaborates on students' profiles in FGDs, retrospective pre-test survey, and in productivity and quality tasks.

#### 3.6.1.2 Students' Profiles

Female students who have successfully completed the course *Computer-assisted Translation* and who are enrolled in the module *Technical and Scientific Translation* were chosen to participate in this study because they were more acquainted with CAT tools, and they were graduating and would be entering the job market shortly afterwards. The *Technical and Scientific Translation* course is mandatory for all students. It focuses on offering the students a general spectrum of technical and scientific texts and the required skills for translating them (Al-Jarf 2017). The module typically consists of a two-hour lecture every week for the period of 15 weeks. In total, seventy final-year undergraduate translation students participated in the study. These students

constitute the two selected cohorts for the current study with 35 student per cohort. All were native Arabic speakers, all females between 21 and 24 years of age. These students did not have any previous experience in MTPE. The homogeneity of participants' profiles justifies for the lack of data weighting in the survey. The following paragraphs expand on the selection criteria for each phase of the study.

**Participants in the Focus Group Discussion.** The participants in the FGDs of this study were 26 female translation students enrolled in the *Technical and Scientific Translation* module in the Spring semester 2019. These students did not take part in the subsequent phases of the study. They were recruited in the semester prior to the one in which the study took place, and therefore they were already graduated by the time the study commenced. The reason why it was important to run the FGDs long prior to the intervention was to (1) maintain enough time to analyse the qualitative data, and to (2) extract details about the students' experiences with computers and MT in order to design the contents of the intervention course that was planned to be given in the following semester. Students participating the FGDs were initially assigned to five groups of an average of five students per discussion group. However, when on site, two groups decided to merge because they wanted to leave earlier for their mid-term exams and the total number of groups was reduced to four with respectively 5, 4, 5, 12 students per group (see 6.5 for further details on how this issue was tackled).

**Participants in the Retrospective Pre-test Survey.** The participants in the post-intervention survey of this study initially involved 31 female translation students who constituted Gb. However, four students opted out from taking the survey as they wanted to leave the classroom once they finished the translation task (see 6.5). Students were not coerced into participating in any part of the research and consent was obtained for each part of the study separately (see Appendix C).

**Participants in the productivity and quality tasks.** The participants in the productivity and quality tasks of this study initially involved 70 female translation students enrolled in the *Technical and Scientific Translation* module in the Autumn semester 2019. Prior to analysis, they were reduced to 60 students upon data cleansing.

## 3.7 The Teaching Intervention

The training programme constitute the base for the independent variable (MTPE training) as it differs from Ga to Gb. In order to maintain the authenticity of the intervention, the teaching of the course *Scientific and Technological Translation* used the same reference book in teaching that is approved by the college when teaching Ga. In Ga, the main reference of skills used is a book titled *Translation as Problems and Solutions: A Textbook for University Students and Trainee*

*Translators* (Ghazala 2012). In Gb, the MTPE training included the following topics (Table 3-3 below) which were modified according to (1) the data obtained in the FGDs about the needs of the students, and (2) the specific translation module objectives in TAUS PE course:

Table 3-3 Topics covered in MTPE training

| Week 1 | Week 2 | Week 3 | Week 4 |
|---|---|---|---|
| • Introduction<br>• Terminology<br>• Why Machine Translation?<br>• Why Post-editing?<br>• Traditional editing vs. Post-editing<br>• Types of post-editing<br>• Post-editing Workflow<br>• خطوات تحرير الترجمة الآلية باللغة العربية<br>• MTPE task- 01 | • Introduction<br>• Recap of basic PE workflow in Arabic<br>• Common MT error patterns in the Arabic Language<br>• Explaining MTPE task - 01<br>• MTPE Task- 02 | • Introduction<br>• Style: Features of the expository text type (النص التفسيري)<br>• Criteria: An investigation of the *Examiner's Marking Sheet* (British Diploma in Translation)<br>• Explaining MTPE Task- 02<br>• MTPE Task- 03 | • Introduction<br>• Organisation and setting up of MT Projects (Preparations for post-editors, post-editing contents with MT engines, and quality control),<br>• Post-editing MT output and evaluating the errors.<br>• Explaining MTPE task - 03<br>• MTPE Task- 04 |

Ideally, for an educational intervention to yield proper results, it should be taught for the whole duration of the semester (e.g., He 2014; Çetiner 2018). However, due to the fact that MTPE training courses for beginners are usually between 6-8 hours of length (e.g., TAUS PE course, TRADOS PE course), I decided to teach the MTPE course for 4 weeks (a total of 8 hrs: 2 hours per week) and give students in both Ga and Gb some reinforcement assignments to be completed in the last part of our meeting in the lab. The reason behind giving these assignments was to compensate for the relatively short training period. In order to encourage students in Ga to complete the extra-curricular assignments and attend the whole period of the course. They had been informed in the welcoming meeting that they would receive an MTPE training course (with college awarded certificate of completion for those who successfully finish it) if they successfully attended the four-week-training and the post-test. (The training course was offered as a two-day workshop for Ga after the study was concluded). In addition, to encourage students in Gb to practice MTPE more and complete the in-lab extra-curricular assignments, they were informed in the welcoming meeting that when they successfully attend the four-week training period, complete the assignments and the take the post-test, that they would receive a certificate of completing a short course in MTPE from the college after the study is concluded.

## 3.8 Translation Quality Assessment

The quality of the product, or the TTs, was defined as a quantifiable variable in section 2.3.1.4. For error type analysis, evaluators' annotations were examined and the total number of times the evaluators mentioned any of the error types (deletion, Aspect of Performance 1 (AoP1), Aspect of Performance 2 (AoP2), or Aspect of Performance 3 (AoP3) of the *Examiners' Mark Sheet*) were identified in the annotations then the average of the total mentions from the evaluators was calculated to run the analysis. Whereas for error count analysis, scores out of 100 were assigned for every student with a standardized TQA model utilised by professional evaluators. The TQA criteria that I decided to use is the *Examiners' Mark Sheet* used in the CIOL Diploma in Translation (DipTrans) (see Figure 3-4 below). It is worth mentioning that in this study, I used the assessment criteria being applied at the time of conducting the study (2019). However, it has changed since then. The reasons why the most recent version of the DipTrans TQA model was not used are because (1) I wanted to use the same criteria used by He (2014) to compare results, and (2) because the criteria used in the current study was the one offered by DipTrans when the gathered TTs were to be assessed.

Figure 3-4 Examiners' Mark Sheet

IoLET LEVEL 7 DIPLOMA IN TRANSLATION: EXAMINERS' MARK SHEET (CONFIDENTIAL – WHEN COMPLETED) *January 2017*

**ALL SECTIONS BELOW TO BE COMPLETED** (one mark sheet per script)

**SECTION A**

| **Candidate Number:** | **Source Language:** | **Target Language:** |
|---|---|---|
| **UNIT** Choose an item. | | |

**SECTION B**

Allocate numerical marks for each Aspect of Performance (AoP)

| **Aspects of Performance** | **Maximum marks** | **Min. Pass mark (60%)** | **Marks obtained** | | |
|---|---|---|---|---|---|
| | | | **Marker** | **Chief Examiner** | **Chief Moderator** |
| **1. Comprehension, Accuracy and Register** | 50 | 30 | | | |
| **2. Grammar, Coherence, Cohesion and Organisation of work** | 35 | 21 | | | |
| **3. Technical Aspects** | 15 | 9 | | | |

| **Final Grade** | **FAIL** ☐ | **PASS** ☐ (60–69) only if **all 3** AoP are passed | **MERIT** ☐ (70–79) only if **all 3** AoP are passed with **no fewer than 35 marks in A1** | **DISTINCTION** ☐ (80–100) only if **all 3** AoP are passed with **no fewer than 40 marks in A1** | **TOTAL:** | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

**SECTION C**

**COMMENTS ON PERFORMANCE**

Complete this section using the Marking Criteria contained in the Marking Guidelines

Please comment on each Aspect separately. Include detailed comments on candidate's performance, and justify the marks awarded

**Aspect 1: Comprehension, Accuracy and Register**

The goal of TQA is to decide on which translation quality is better. But *how* better depends on many factors that differ from one TQA model to another. Indeed, it was argued that traditional TQA practices suffer from too much focus on the end-product (Suojanen et al. 2014). However, it is the aim of the current study which is evaluating the effectiveness of a training intervention after its completion that justifies the focus of the current study's TQA on the end-product as analytical assessment lies outside the scope of this research. The purpose of the current study requires a TQA model that is consistent, easier to learn and practice by the evaluators, objective, reliable, timesaving, cost-efficient, and one that can assess both HT as well as MTPE.

Since the goal behind measuring the quality of TT is as objectively as possible to compare the scores of students in both cohorts and to decide on whether they produced comparable scores while identifying the error types in TTs produced by both translation methods, the *Examiners' Mark Sheet* that was used in the DipTrans programme in 2019 was found to meet the criteria required for the scoring of the TTs within the scope of this study.

According to the DipTrans main page ( https://www.ciol.org.uk/DipTrans ) (accessed: 2 December 2019), exam formats that can be corrected using this marking criteria include written translation of general texts, and written translation of a semi-specialised text in one of the following domains: Technology, Business, Literature, Science, Social science, and Law. In addition, one of the pros of this marking criteria is that it was designed to be used to correct a range of language combinations including the language combination English-Arabic.

The overall assessment criteria of the *Examiners' Mark Sheet* used for error analysis requires that in order for a translator to pass the translation task, that deletion of parts from the ST in the TT must not exceed 5%, and that translators must pass all three Aspects of Performance (AoP) for the task. They are:

- **AoP1 (Comprehension accuracy and register):** The correct transfer of information and evidence of complete comprehension and appropriateness of rendering and lexis.
- **AoP2 (Grammar, cohesion, coherence and organization of work).**
- **AoP3 (Technical aspects):** Relating to punctuation, spelling, accentuation, transfer of names, dates, figures, etc.

To mark these three aspects, a proportion of 50, 35 and 15 is respectively allocated for the three AoPs (Figure 3-4 above). In addition, the criteria are accompanied by detailed guidelines on how to mark each AoP and the total scores are allocated specific cells in the Mark Sheet table.

#### 3.8.1.1 Error Type

Error type in the current study refers to deletion of 5% or more plus the three Aspects of Performance (AoP) illustrated in the *Examiners' Mark Sheet*. Namely, (1) comprehension accuracy and register, (2) grammar, cohesion, coherence and organization of work, and (3) Technical aspects. These AoP in addition to deletion were used as the method of correction in this study. This method was used to help answer the first sub research question or RQ3 in which the most common errors in HT and MTPE are identified and calculated. To do so, the evaluators were asked to use the AoPs mentioned in the *Examiners' Mark Sheet* as a reference for the types of errors then they were requested to provide an annotation in which they comment on their first personal overall opinion about the translation and write about the most common errors they encountered in each TT they reviewed. The evaluators checked both the ST and the TT for each translation task that they evaluated.

#### 3.8.1.2 Evaluators

When evaluators are needed for a translation study, there are certain criteria to take into consideration so that the decision of recruiting them is easier for the researcher and the results are more reliable. According to Castilho et al. (2018), such criteria include: the nature of the task required to be undertaken by the evaluators, i.e., whether a bilingual evaluator is necessary (if the evaluator is required to assess based on both the ST and the TT) or is a monolingual one sufficient (in the case of assessing the fluency of the TTs for instance). In addition, there is the question whether the evaluator has to be professional and experienced with the TQA method used, which might cost more, or will amateur evaluators suffice and be more cost-effective especially in studies of a larger sample size. Finally, there is the decision of recruiting an individual or a group of evaluators based on the nature of the task required.

The texts used in the current study are of a technical nature, which set the first criterion which is to search for an evaluator who preferably has experience of assessment of such text types. In addition, when deciding on whether to recruit amateur or professional evaluators, the number of TTs had the largest weight in the process of deciding. Since the number of texts is relatively high (60 texts), I had to search for evaluators who preferably have some sort of experience with TQA, but they were not necessarily required to know how to use the TQA model used in the current study beforehand. Therefore, to make up for the evaluators' inexperience with this specific TQA model, I had to consider recruiting more than one evaluator so that the assessment is conducted in a way that allows for an averaging of their scores to moderate their strong negative or positive personal biases (Castilho et al. 2018)

Thus, in order to maintain the reliability of the test results and to reduce bias, two English-Arabic translation teachers in higher education (one is in a university in Saudi Arabia, and the second is in a university in the UK) who have taught and assessed technical translation were recruited and they received money to assess students' performance in both Ta and Tb. Both evaluators had 10 years or more of experience in teaching and evaluating translations at undergraduate levels. To ensure the that the evaluators were not affected by the method of translation used, students' IDs were concealed, and e-files of TTs were only numbered from 1 to 120 (60 files of Ta plus 60 files of Tb) without any indication of the method of translation used to generate the TT. The evaluators were first acquainted with the contents and guidelines of the *Examiners' Mark Sheet*, they worked together through online meetings to negotiate annotations and error marking. During the marking process, the two evaluators analysed the errors based on the *Examiners' Mark Sheet* framework and they reached a consensus after their discussions. Although it was agreed that they refer to a third evaluator (the researcher) if any disagreement arose, they did not report any disagreement in the marking process.

#### 3.8.1.3 Inter-rater Reliability

To increase the credibility of results in studies that involve more than one evaluator, researchers in statistics use different tests and formulas to compute the inter-rater reliability (IRR). According to Graham et al., there is a substantial difference between inter-rater reliability and inter-rater agreement where inter-rater reliability is 'the measurement of the consistency between evaluators in the ordering or relative standing of performance ratings, regardless of the absolute value of each evaluator's rating. (2012: 5). On the other hand, the authors define inter-rater agreement as 'the degree to which two or more evaluators using the same rating scale give the same rating to an identical observable situation (e.g., a lesson, a video, or a set of documents). (Graham et al. 2012: 5). The importance of knowing the difference between inter-rater reliability and agreement stems from the reasons behind evaluations made in educational settings where two or more raters are conducting the evaluation. In some cases such as the current study, the consistency of raters' judgments about the relative levels of performance is the aim of running the evaluation. Therefore, inter-rater reliability is the test I decided to calculate in order to address the issue of consistency within the rating system implemented in my study.

IRR can be computed through a number of statistics. Some of the most common indexes are the percentage agreement, kappa, product-moment correlation, and interclass correlation coefficient (Lange 2011). If the statistical test of IRR is high, it indicates a high degree of agreement between raters whereas low IRR refers to low agreement between the raters (whether they are two or more) (Lange 2011).

IRR in translation quality studies have revealed low agreement between raters in a number of studies (e.g., Carl et al. 2011; Jia et al. 2019; Vieira 2016). Given that evaluating translation quality is mainly subjective and more complicated when more than one evaluator is assessing the tasks, it has been accepted in the field that the IRR is slightly above chance (when using kappa) or when it is above 75% when using percent reliability (Graham et al. 2012).

Choosing the index to test inter-rater reliability depends on a number of factors such as the number of raters and the number of rating levels. The current study recruited two evaluators to examine four types of error in translation tasks. Therefore, calculating the percentage agreement is considered suitable. The percent agreement for two raters is calculated through counting the number of ratings in exact agreement, counting the total number of ratings, then dividing the total by the number of exact agreements. The resulting number represents the percentage of agreement between the two raters. If the agreement percentage is lower than 75%, raters are advised to discuss their nonagreements and reproduce results of higher percentage of agreement (Graham et al. 2012).

As a general rule in percentage of agreement index, any percentage lower than 60% is considered unacceptable, between 60% and 75% is considered acceptable, and between 75% and 90% is considered good (Lange 2011). In the present study, the first time IRR was tested, the result was 68% (82 exact agreement out of 120 total ratings). The evaluators discussed their nonagreements without my interference as a third evaluator, and their second IRR result improved to 76% (91 exact agreement out of 120 total ratings).

#### 3.8.1.4 Error Count

Error count in the current study refers to the total score after deducting the total count of errors from a hundred based on the criteria set in the *Examiners' Mark Sheet* of the DipTrans. The total remaining after all points for error are deducted is the total score of each student. This method was used to help answer the second sub research question of RQ3 in which a comparison between quality scores resulting from HT and MTPE is made. The marking criteria used for counting errors was as follows: one point was taken for every AoP error spotted by the evaluators. For example, if the evaluators found two types of errors in one sentence such as an error in accuracy and an error in technical aspects, two points (one point from AoP1 and one point from AoP3) will be deducted. However, the marking criteria also state that translations with 5% or more of source text missing will be automatically awarded a Fail mark. I asked that this criterium is also applied by the evaluators for comparison of deletion between HT TTs and MTPE TTs.

The ranking categories mentioned in the marking criteria (i.e., distinction and merit) were intentionally neglected as they fall outside of the scope of the study. The scope of the study focuses on the total score achieved by the student in each aspect of performance, whether the student failed or passed, and on the total score out of 100 allocated for each student.

## 3.9 Data Collection

Data for this study were collected using multiple sources and methods, these are further discussed in the following corresponding subsections. The alignment of the time when data were collected with the activities of data collection and data analysis methods are outlined in Table 3-4 below.

Table 3-4 Processes of collecting data

| Order | Date | Action | Activity |
|---|---|---|---|
| **Phase 1** | April 2019 | Students' Opinions | 4 FGDs (total of 26 students) |
| **Phase 2** | October-November 2019 | Ta (productivity and quality) | 60 HT TTs (29 Ga and 31 Gb) |
| | | Teaching Intervention | 29 students (Ga): 4 weeks HT training<br>31 students (Gb): 4 weeks MTPE training |
| **Phase 3** | November 2019 | Tb (productivity and quality) | 29 HT TTs (Ga) and 31 MTPE TTs (Gb) |
| | | Students' Opinions | 27 survey responses (Gb) |

The following sections describe the data collection processes chronologically.

### 3.9.1 Focus Group Discussions

Focus group discussions were conducted in April 2019 before the intervention began to explore the opinions of translation students towards MTPE and to feed into the design of the intervention (the intervention took place in October 2019). The participants in the FGDs were different from those who participated in the intervention and tasks. The reason for recruiting a different group of participants for the FGDs is because the qualitative data collection was one semester prior to the intervention. I wanted to explore the opinions of students representing the same level that was targeted in the study. The aim of the FGDs was to gather as much qualitative data about students' opinions and feelings towards MTPE as possible. Another reason was to have sufficient

time to analyze the qualitative data and design the MTPE course contents based on the average level of MT experience of the students. A semi-structured interview approach was selected for the discussions. This approach provides an opportunity for the interviewer to ask follow-up questions that emerge during the interviews (Cohen et al. 2013; Onwuegbuzie et al. 2009). All discussions were conducted in Arabic and in private meeting rooms at the university. Each discussion lasted for about an hour. In addition, each FGD constituted of 4 to 5 students (Morgan 1993; Guest et al. 2017). The main questions are listed in Appendix B.

Table 3-5 below describes the time of data collection, the total number of results obtained and the number of students in each focus group.

Table 3-5 Summary of data collection- Opinions

| Data | Collected in | Total Results | Result Details |
|---|---|---|---|
| Focus Group Discussions | April 2019 | 26 students (data corpus) | FGD1: 5 students |
| | | | FGD2: 4 students |
| | | | FGD3: 5 students |
| | | | FGD4: 12 students |
| Retrospective Pre-test Survey | November 2019 | Experiment:<br>Gb 27 students | N/A |

### 3.9.2 The Process of Data Collection

The intervention that was prepared to be given to the experimental group involved 4 weeks of teaching plus 4 activities. Before the pre-test that was prepared for both groups and before the students in the experimental group began engaging in the MTPE training course, two ice-breaking sessions were performed (one session for each group) to allow me to introduce the research, explain the idea behind it, and for the students and me to get to know each other. In this welcoming session for the control group, I explained to them that I would be teaching them the same course (Technical Translation) in the same conventional way they are used to but through the use of different texts. The only changes are that they will be attending the lessons in the computer lab and that they will be using computers to type their translation (as they usually translate using pens and papers), and that they will take a pre-test and a post-test in which they will translate using the conventional HT method. On the other hand, in the welcoming session for the experimental group, I informed the students that they will be training to translate the same

text types as the control group but through post-editing MT output and that the training sessions would take place in the translation lab where they would be using computers throughout the training session.

Focusing on *Scientific and Technological Translation*, the two cohorts of students took a baseline test (Ta) before the commencing the four-week teaching intervention in order to define their translation ability at the starting point and to provide a reference for comparison against the post-test (Tb) results. Afterwards, they underwent a four-week period of translation training, during which Gb was trained to translate using the MTPE method with the assistance of Google Translate, while Ga was trained to translate the same texts using the conventional method (HT). At the conclusion of the four weeks, both groups of students took part in Tb, also dealing with the translation of texts of technical nature, so as to gauge their translation ability after the training. Both tests (Ta and Tb) generated a number of test results. Comparisons were done between the two groups' translation productivity and quality, and conclusions were reached in terms of whether Gb have made more, similar to, or less improvement than Ga, and whether any difference in their improvement (if any) can be said to be significant enough to provide evidence that MTPE training was effective.

During the intervention period, I aimed at maintaining the difference between the two cohorts in that students in Gb were trained to post-edit MT generated texts using GT, while students in Ga received the conventional HT instruction of the course. Except for this difference, Ga and Gb shared the same level of training and took the same number of assignments during the course of their training. They both studied the modules of the same translation course by the same teacher (the researcher).

### 3.9.3 Translation Productivity

In November 2019, students received translation e-briefs that preceded the tasks explaining all the steps required for both HT and MTPE. Each brief included instructions on how to write their IDs, instructions on operating the timer, how to carry out the task, instructions asking them to produce the best quality possible, and the type of dictionary allowed for use during the task.

Students in Ga received an e-document that includes the e-brief and the source text. In an attempt to maintain similar variables and for the sake of the authenticity of the task environment, the students were instructed not to use an online dictionary and they were allowed to use Oxford paper-based dictionary in the assignments and in Ta and Tb tests, and they typed in their translations on a Microsoft Word document. Students in Gb received the same e-document that included the e-brief and the source text, but they were instructed to use the new method of

translation that they learnt, that is to generate a MT output using Google Translate then fully post-edit the output to the best possible quality. Students in Gb also used Microsoft Word to paste the MT output and post-edit the texts from English into Arabic. In order to maintain variables as controlled as possible, they were instructed not use online dictionaries and if they needed one, they were advised to use the same dictionary used by students in Ga.

Similar to Green et al. (2013), students in this study were under time pressure as they were instructed to complete the task within 60 minutes (however, those who did not finish translating were not instructed to stop by the end of the 60-minute period). The total time spent on the tasks was self-reported by students through means of built-in timers in the computers which they were requested to activate before commencing the HT or the MTPE tasks. However, this method of calculating time (which was used by Lee and Liao 2011) could be unreliable if each student paused and reactivated it while the purpose of the whole task was to re-create an authentic translation test scenario. However, even if a more sophisticated method of calculation was to be used, threats to the reliability of measurement still exist. As it is possible that students who use PE might gain speed over the course of time of the task due to gained familiarity. Also, students in both groups (HT and MTPE) might lose speed due to fatigue or boredom. In an attempt to deal with the issues arising from using a primal measurement technique, students were instructed not to pause the timer at any given moment while they were translating. They were advised to only stop the timer and record the finishing time that appeared on the clock when they were confident that the task had been completed and did not require further changes. Reasons for not being able to use a more reliable method for calculating the total translation time are listed in the limitations (see section 6.5).

### 3.9.4 Translation Quality

The second set of data (experiment on translation quality) were collected in November 2019. The Error type data generated 62 TTs from students in Gb (31 Ta and 31 Tb). On the other hand, error count produced 120 test results which were generated by 29 students from Ga and 31 students from Gb, both cohorts having finished the pre-test (Ta) and the post-test (Tb).

As for error type data, I prepared the annotation experiment by teaming up the evaluators and demonstrating to them that the error metric for annotations is based on the aspects of performance listed on the *Examiners' Marker Sheet* in addition to their remarks on deletion if it resulted in missing 5% of more of the ST in the TT. Evaluators were asked to provide one annotation per TT in which they highlighted the most reoccurring errors based on their first reading of the TT.

As for error count data, the e-brief had instructions for students in Ga to type in their translations into an MS Word document and email it to my university email at the end of the experiment. Students in Gb, on the other hand, were instructed to use GT to generate an Arabic MT output that they then copied and pasted into an MS Word document. After that, the e-brief illustrated that students should apply the PE skills they learned in the teaching intervention to generate a publishable TT before emailing it to my university email.

The initial assumption was that, after four weeks of training, students in Gb would have improved quality when compared to students in Ga, because they had been trained with access to MTPE training course. The difference of improvement between the two groups is shown by the score difference found in the students' results obtained in Ta and Tb. It was assumed that the scores of Tb would be higher than the scores of Ta for both Ga and Gb, because students may have improved in terms of translation ability after the translation course regardless of the method of translation used in the training. A standard improvement should be evident when the scores obtained in Tb are contrasted with those reached in the previous Ta. Here, the assumption is that all students will get a higher score in Tb than in Ta, primarily because based on the carefully selected texts for Ta and Tb (see section 3.5.1.1), the difficulties encountered in translating Tb were similar to those encountered when translating Ta, and all the students from the two cohorts have had additional four weeks of training. However, the important difference lies in finding out whether Gb scored as well as Ga in terms of quality. The improvement is shown by the score difference between Ta and Tb, namely, Tb-Ta. The assumption was that Gb scores would show more/similar improvement when compared to Ga scores because the students in Gb have been trained with access to MTPE strategies while students in Ga have not. If we put this assumption in the form of numeral scores, it would mean that the score of Gb (Tb-Ta) should be higher or equal to that of Ga (Tb-Ta).

### 3.9.5 Retrospective Pre-test Survey

Right after the conclusion of Tb, students in Gb were asked to fill in electronic surveys they were provided with then email them to my university email before leaving the language lab. The initial number of students in Gb was 31 students. However, I only received 27 responses at the end of the experiment. Reasons for this disparity in numbers is provided in section 6.4.

Retrospective pre-test surveys are used as an instrument to measure at the same interval of time (which is usually after the intervention) the degree and change of participants' opinions before and after an intervention (Chang and Little 2018; Gouldthrope and Isreal 2013). Retrospective pre-test surveys were first introduced by Campbell and Stanley in 1963 (Thomas et al. 2019). The

instrument in the current study includes 7 statements with multiple choices for each statement. Each statement has a before and an after section to which the student responds simultaneously. Some statements have 2 choices, and some have more. The statements for the current research were derived from the post-test questionnaire created by Daems (2016) who presented them to the participants in the form of a fill-in-the-blank questionnaire. Using the same statements was intentional as it would allow for comparison against Daems' findings.

Several measures were taken to ensure the survey was valid and as free from errors as possible. Firstly, face validity of the survey was established. All seven statements in the survey were assessed by two experts in the field of translation who are familiar with the topic I am researching. They were asked to ensure the statements were not leading or confusing when read by the students. Secondly, survey piloting was conducted. Although it was not feasible to pilot the survey with translation students because I was in the UK when I had it piloted, four of my colleagues completed the survey and used a think-aloud protocol (Ericsson and Simon 1998) to ensure that translation students' interpretation of the survey elements would be similar to my intention. Each one of the four volunteers was asked to read the statements of the survey and provide their interpretation to me. Additionally, the survey was translated into Arabic to guarantee students' full understanding of the statements, and I sought the assistance of two of my colleagues who have got high proficiency in both languages to check the English version of the survey and the accuracy of its Arabic translation.

### 3.9.6 Piloting the Study

In the original plan for data collection, piloting the discussion questions and the survey with students was an essential part. However, due to the short period of time allocated for the data collection trip (data were collected in two trips to Saudi Arabia; trip one was one-week long, and trip two was 5 weeks long), I could not recruit students to pilot the focus group discussions and the survey. But in order to overcome this issue, I piloted the discussion questions with two of my translation colleagues, and the survey statements with four of my colleagues in order the make sure that the discussion questions and the survey statements were clear and that what the participants understood from the survey exactly what I meant. In addition, both the focus group discussion questions and survey statements were discussed and approved in a college committee meeting in the college where this study took place before I was granted the approval to start data collection.

## 3.10 Data Analysis

Table 3-6 below lists the RQs, their justifications, and the unit of analysis used for every type of collected data.

Table 3-6 Research questions' justifications and units of analysis

| No. | Research Question | Justification | Unit of Analysis |
|---|---|---|---|
| RQ1 | **Opinions:**<br>What are the differences in students' opinions about HT and MTPE?<br>a. How rewarding is PE compared to HT?<br>b. How useful is MT output according to translation students?<br>c. Which translation method is perceived as being faster?<br>d. How is the quality of both methods of translation perceived?<br>e. Which translation method is the most preferred translation method? | The answer to this question will uncover opinions and perceptions about MTPE before the teaching intervention | Group data |
| RQ1 | f. Is there a difference in perception before and after the intervention/experiment? | The answer to this question (in addition to the previous Qs a. through e.) will examine the differences in students' opinions about MTPE after the intervention. | Survey responses |
| RQ2 | **Productivity:**<br>What are the differences in the process between HT and MTPE? | The answer will provide evidence of whether MTPE achieved more productivity than HT. | Group mean total time in mins |
| RQ3 | **Quality:**<br>What are the differences in the product between HT and MTPE? | The answer will provide evidence of reoccurring errors in Arabic MTPE TTs.<br><br>The answer will provide evidence of whether MTPE TTs achieved comparable scores to HT. | **Error Type**: Total number of error types (annotated by evaluators)<br><br>**Error Count**: Group mean total score out of 100 |

### 3.10.1 Qualitative Data Analysis

For the qualitative data, thematic analysis of data from the FGDs was selected. Thematic analysis is the processes of working with raw data in order to identify and interpret the key themes and ideas (Saldanha and O'Brien 2014). The thematic analysis in the current study followed a deductive approach that involves analysing data based on a structure predetermined by the research question. Therefore, I used the research sub-questions as a guide for grouping and analysing the data. One of the main aims of this study was to explore and compare students' attitudes towards MTPE as a part of a larger-scale study, and the advantage of the deductive thematic analysis approach is that it is both quick and easy particularly because the questions in the focus group discussions were based on the sub-questions of RQ1, which means that I may be able to predict some of the likely responses from the students. Thematic analysis depends on collecting and organizing data followed by generating categories and themes, which in this case were already predetermined. After data were coded, emerging understandings of data are tested and alternative explanations of the data are searched for, and finally the analysis report is written (Fox 2004).

#### 3.10.1.1 The Process of Thematic Analysis Using Microsoft Excel

The thematic analysis was intended to be conducted using Nvivo. However, upon transferring the students' transcripts to Nvivo, all Arabic texts which read from right to left were reversed to read from left to right and the analysis was not possible. The actual thematic analysis was conducted through the use of Microsoft Excel. In this study, I used the Robinson (2021) methodology that goes as follows:

- Deep immersion in the data
- Generating initial codes and themes
- Tabulating themes against data segments
- Exploring theme frequencies
- Producing the report

**Deep immersion in data-** Transcripts of every question of the FGDs were transferred from a Microsoft Word document that includes all the responses to a password-protected Microsoft Excel spreadsheet. The Excel file was composed of six sheets with every sheet representing the comments/answers of one focus group question. Following is the order of the Excel sheets in the file:

Sheet 1: MT Experience

Sheet 2: Degree of Reward

Sheet 3: Usefulness

Sheet 4: Speed

Sheet 5: Quality

Sheet 6: Preference

The first column included a student identifier, and the second column included each student's response in a separate cell. I repeatedly read the data carefully and slowly, colour-coding certain words and expressions in responses for possible codes and added other initial ideas in a third column that was named *initial notes* (Robinson 2021). This step was completed through ensuring immersing reading of the entire dataset until I felt I was strongly familiar with all the data and patterns started to appear.

**Generating initial codes and themes-** in this step of the process, I added a fourth column that was intended to contain *names of codes.* Based on repeated reading, terms and words that describe content in students' responses were added in this column. Through this, common patterns, words and ideas, which are the common goal of a thematic analysis, appeared.

**Tabulating themes against data segments-** in this step, I attached data segments to themes in a tabulated form. According to Robinson (2021), this step provides foundation for agreement-checking and enables frequency calculation processes.

**Exploring theme frequencies-** A frequency calculation cell was added at the bottom of each column to provide accurate numbers about the prevalence of themes when writing up the report in the final step.

**Producing the report-** Due to the design of the study being mixed-method, numerical and textual data about students' responses were integrated and presented in the report.

This analysis approach was cost-effective in which a widely used application software suit was used for organizing, coding, and classifying data. Particularly, it used the colour-coding and sorting features of Microsoft Excel without requiring any advanced knowledge of the software.

### 3.10.2 Quantitative Data Analysis

Quantitative data analysis was used for the analysis of the opinion retrospective-pre-test surveys, the comparison of translation productivity, and the comparison of translation quality (both error type and error count). Following are descriptions of how data analysis was run for each type.

#### 3.10.2.1 Retrospective Pre-test Surveys

For the quantitative data obtained from the retrospective pre-test survey, a paired-sample *t*-test was conducted to examine the change of opinion within the experimental group using percentages. Since the answers are defined based on an ordinal scale, the Wilcoxon test (non-parametric approach) for two paired samples was used to test whether the participants opinions about MTPE after the intervention significantly changed from before the intervention. The test was applied to each question about opinions. In addition, The Z-score was calculated to provide an interpretation of the effect size. The z-score is a numerical measurement that describes the relationship of a certain value to the mean of a group of values and is measured in terms of SD from the mean (Hayes 2020). When the Z-score is positive, it indicates that the raw score is higher than the mean average, and when the Z-score is negative, it reveals that the raw score is below the mean average. To analyze significant differences, I considered calculating confidence intervals, effect size, and p- values. I used the statistical software SPSS to calculate the data.

#### 3.10.2.2 Translation Productivity

For the quantitative data obtained from measuring total translation time, I ran two types of comparison: longitudinal and horizontal. The two longitudinal comparisons aimed to calculate the actual productivity differences between Ta and Tb for both groups, so as to confirm whether students in Ga and Gb have made any improvements between the two tests. I conducted paired-sample *t*-tests to the examine the productivity gain in each group before and after the intervention. The effect size was measured through examining the p-value. On the other hand, the horizontal comparison was carried out by considering the productivity of both groups (Ga and Gb) while considering both time intervals (Ta and Tb). The horizontal comparison aimed at finding out whether MTPE has played a role as a factor in affecting the productivity of the experimental group (Gb). Repeated measure ANOVA was conducted for this analysis, and the effect size was measured through examining eta square ($\eta^2_p$).

#### 3.10.2.3 Translation Quality

In an attempt to answer RQ3 *Is there a difference in the overall quality between the product of HT and the product of MTPE*, I ran two types of analysis, for the quantitative data obtained from error type to answer sub-research question 3.1, I counted the total number of times each AoP of the *Examiners' Mark Sheet* was mentioned (hereafter, Number of Mentions) by the evaluators in their annotations for Gb. The reason for limiting the calculations to Gb only is that the sub RQ seeks to explore the identified errors in both HT and MTPE which were used only by students in Gb as

students in Ga used HT in both Ta and Tb. Then I ran a comparison between the Number of Mentions in Ta (HT) and Tb (MTPE).

On the other hand, to answer sub-research question 3.2, and to successfully run a quantitative analysis of error count, translation quality was operationalised in section 2.3.1.4 into a measurable variable which is a total score of 100. A comparison of translation quality scores was conducted through comparing the mean score of each group. The translation tasks in both groups (Ga and Gb) were first marked based on the criteria set out by the *Examiners' Mark Sheet* (see Figure 3-4 above). Each student's task was awarded a score out of 100, and then the value of all the scores was entered in the marking sheet for the means to be calculated.

For the quantitative data obtained from error count to answer sub-research question 3.1, I ran three types of analysis: horizontal Ta comparison between Ga and Gb, Horizontal Tb between Ga and Gb, and longitudinal Gb comparison between Ta and Tb. Firstly, the horizontal analysis of Ta for both Ga and Gb was run to calculate the baseline scores through measuring the group mean. Baseline scores (group mean) were calculated for each group (Ga and Gb). Group means were calculated by adding the total scores of the whole group, then dividing the total scores by the number of students in the group (see example Figure 3-5 below).

Figure 3-5 Calculating the group mean- Example

| Group | Student 1 score | Student 2 score | Student 3 score | Student 4 score | Total score | mean |
|---|---|---|---|---|---|---|
| Ga | 80 | 75 | 87 | 79 | 321 | 321/4 = **80.25** |
| Gb | 86 | 78 | 88 | 75 | 327 | 327/4 = **81.75** |

Secondly, after the baseline scores were calculated and entered in an excel sheet for comparison, the second horizontal analysis of Tb for both Ga and Gb was run through conducting repeated measures ANOVA, and the effect size was measured through examining eta square ($\eta^2_p$). finally, longitudinal analysis of the results from Gb in Ta and Tb was conducted using paired-samples *t*-test and the effect size was measured through examining the p-value.

#### 3.10.2.4 Data Cleansing

The quantitative data collection and analysis in the current study were conducted to explore the effectiveness of MTPE in the productivity and quality of translation among undergraduate translation students. The study was initiated with 35 students in each group. However, upon data-

cleansing stage, outliers were discovered in both groups. Students have either spent too much or too little time working on the task. Because the scope of the study did not include criteria allocated for issues related to learning differences (such as dyslexia), I removed 6 entries from Ga and 4 entries from Gb to facilitate the interpretation of data and the study was carried out with 29 students in Ga and 31 students in Gb. Then normality distribution was tested again.

Before conducting the paired-samples *t*-tests, the assumptions of normality were tested, and outliers were removed as previously discussed. According to the Shapiro-Wilk test which is highly recommended for small sample size (Ghasemi and Zahediasl 2012), the distribution of the time spent on completing the tasks in both the pre-test (.387) and the post-test (.559) was approximately normal (the value is greater than 0.05. Figures 3-6 (Ta) and 3-7 (Tb) below demonstrate the normal distribution curve for both groups).

Figure 3-6 Normal distribution curve of Ta

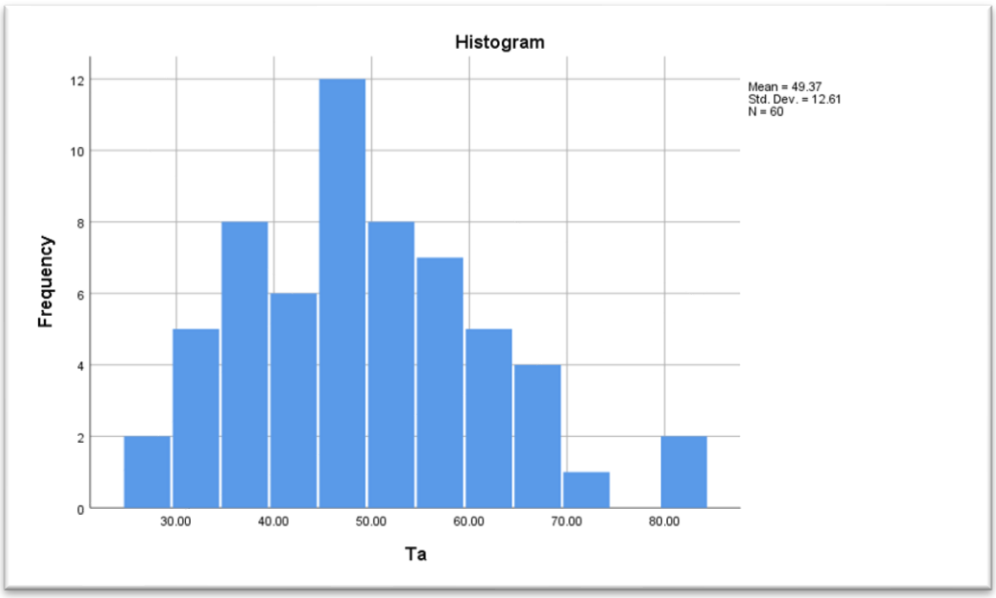
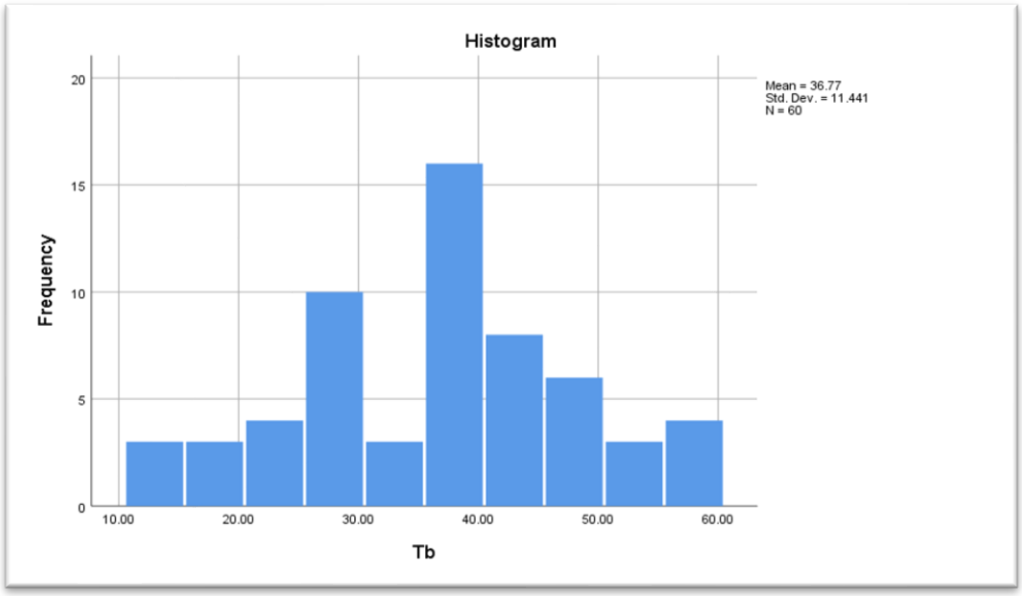



Figure 3-7 Normal distribution curve of Tb




## 3.11 Ethical Considerations

Ethical considerations are indeed an essential aspect at all stages of the research process (Miller and Brewer 2003). Different parts of the study shall be ethically considered from the design of the study and how participants are recruited, to the treatment they received during the study, to consequence of their participation (Miller and Brewer 2003). The main aim of research ethics is to

ensure that participants will not be harmed in any shape at any stage of their participation (Cohen et al. 2013). Therefore, this study considered all ethical issues that may arise throughout the conducting and reporting of its different stages.

In this regard, an ethical approval was sought and guaranteed from the University of Southampton before commencing the study (See appendix A). Moreover, an approval to collect data was obtained from the College of Languages and Translation, King Saud University (See Appendix G).

All participants were informed about the purpose of the study, and that their participation in it was voluntary and that they were completely free to withdraw from participating at any stage of the study without any further responsibilities. Each participant received a Participant Information Sheet (PIS) that detailed all the previous information followed by a consent form that they signed as formal evidence of their agreement to take part in the study (See appendices- Appendix C and Appendix D). Both the PIS and the consent form were translated into Arabic to ensure participants understood their rights and the degree of their involvement in the study (See appendices- Appendix E and Appendix F). In addition, participants were informed that their participation in the study would not be part of their academic assessment and that their comments and responses would not be identified by anyone other than the researcher. It was clarified in the PIS that their identities would not be disclosed when reporting the results of the study and that only the researcher involved in this study would have access to their information. Rather, pseudonyms were assigned for students since the beginning of the study and the students used their new IDs in every part of the study. The following measures were considered when setting up the study: in order to protect students' privacy, they were named from Number 1 to Number 31 (abbreviated as N1, N2, N3, etc.), so as to keep students' identities confidential. Therefore, for an easier reference, students will be referred to as GaN1 (i.e., student number 1 in the control group) or GbN15 (i.e., student number 15 in the experimental group) and so forth. The study did not involve collecting any sensitive data from the participants.

## 3.12 Reflection on Researcher Positionality

Researcher positionality refers to the position of the researcher with respect to others in the study and his/her influence on the research (Hammond and Willington 2012). In all three phases of the current study, participants are considered the main providers of data, whereas I am an insider-researcher in the college, the data collector, and the data analyser of their views on MTPE, translation productivity and quality scores. Thus, this section discusses my positionality in the study.

Scholars suggest that when conducting research, the researcher is to take different roles. These roles can vary from being a member of the group being researched (i.e., an insider) to being a stranger to the group being investigated (Unluer 2012). In this study, I considered myself an insider-researcher as I have worked as a translation lecturer at the College of Languages and Translation where this study is taking place. In addition, I used to teach the course *Computer-assisted Translation*, which has enabled me to gain experience in contacting students. Being an insider-researcher has many advantages. Since I have worked in this college for over 15 years, I have a strong understanding of the cultural and educational practices of the context of the study as well as the rules and regulations of the college. As a result of this understanding, I was capable of seeking the help of my colleagues in the recruitment of participants and approaching the students in this study. Also, being an insider-researcher played a significant role in facilitating the process of data collection as I speak the same mother tongue as the students, and I am aware of the local rules and values that govern their interactions. According to Coloma (2008), it is hard to gain the trust of participants if the researcher does not share any of the key features of the participants, such as their gender, language, or educational level. This is why sharing one or more of these characteristics facilitated my interaction with the students involved in the study while maintaining the trust and the high level of understanding during the data collection phase.

However, recognising the negative influence of my position on the research, I tried to separate myself from my personal views and beliefs. For instance, while conducting the focus group discussions, I maintained my silence and only asked questions of clarifications without assuming that I knew what the participant meant. This has helped reduce the researcher influence on participants' views or the level of their interaction. It should be noted that I returned to the college as a researcher and not as a lecturer, which means that I had the benefit of the participants knowing me but without the worry that I was responsible for any of their academic assessments. All participants in the study knew in advance that the optional course that I offered and taught was for research-related purposes only.

## 3.13 Summary

This chapter discussed the pragmatic approach in addition to the experimental and mixed-methods design, as well as the methods used in the current study. A discussion of the study settings, sampling, piloting, ethical considerations and researcher positionality have also been provided and explained. The upcoming two chapters will examine in depth the results of the study (Chapter 4) and the discussion of those results (Chapter 5).

# Chapter 4 Results

## 4.1 Introduction

In the previous chapter, an overview of the research methodology and the instruments that have been used to conduct the study were provided. In this chapter, I present the results of the analysis of the collected data. Firstly, the transcriptions of the focus group discussions and the responses of the retrospective pre-test survey to answer RQ1 about students' opinions are reported. Secondly, the recorded task completion time in minutes to answer RQ2 about productivity is outlined. Finally, the total scores in the translation task and the identified errors in translation to answer RQ3 about translation quality are listed. Thus, the second section will be comprised of descriptive as well as quantitative results, and the subsequent sections will focus on the quantitative analysis of data.

This chapter comprises four sections (including this introductory section). Section 4.2 encompasses results that tackle RQ1 about students' opinions. Under this section, section 4.2.1 provides the findings from the qualitative analysis of students' responses in the focus group discussions (FGDs), whereas section 4.2.2 presents results drawn from responses to the retrospective pre-test survey. In section 4.3, comparisons of task periods are made and results that tackle RQ2 about productivity are presented. Finally, section 4.4 tries to answer RQ3 through error type (section 4.4.1) and error count (4.4.2). In this chapter, I illustrate the significant results from the collected data; in chapter 5 these results will be examined and discussed in detail.

Before presenting the results, the following list of abbreviations and their meaning is provided in Table 4-1 in order to facilitate the reading of the coming chapters (I have adapted the list created by He (2014) and modified it to suit this study):

Table 4-1 List of abbreviations

| | |
|---|---|
| Ga | Group A, the control group, female translation students who used HT in both Ta and Tb |
| Gb | Group B, the experimental group, female translation students who used HT in Ta and MTPE in Tb |
| Ta | The pre-test, the baseline test, students in both Ga and Gb used HT in this test |
| Tb | The post-test, students in Ga used HT while students in Gb used MTPE in this test |
| fs | Final score: the overall score of the test out of 100 |
| AoP1 | Aspect of Performance no. 1: Comprehension accuracy and register |
| AoP2 | Aspect of Performance no. 2: Grammar, cohesion, coherence and organization of work |
| AoP3 | Aspect of Performance no. 3: Technical aspects |

## 4.2 Opinion Results

In this section, I present the analysis of two instruments, which are the data derived from the focus group discussions, and the responses from the retrospective pre-test survey. Therefore, to answer RQ1, i.e., what are the differences in students' opinions towards HT and PE, I have two sections: What the responses from the focus group discussions informed us, and I offer examples as the analysis requires. In addition, this section includes what the responses from the retrospective pre-test survey revealed.

### 4.2.1 Findings from Focus Group Discussions (Pre)

This section illustrates the input from translation students with regard to their opinions about MT and MTPE in the FGDs. The FGDs included questions that constituted predetermined codes derived from existing literature about translators' opinions towards MTPE (Daems 2016), i.e., degree of reward, usefulness, translation speed, translation quality, and personal preferences of translation method. I only added a question about MT use and provided a justification for why it was added.

To accomplish the specific objective of RQ1, 26 students were asked to discuss their opinions regarding every question. Although this part of the study is largely guided by a qualitative approach, quantitative data can also be valuable to enhance the analysis. As a result, there will be some quantitative and frequency discussion. The reason behind the frequency calculation is to compare students' responses in the FGDs with the responses from the survey. Therefore, the numbers of similar responses from students were mentioned before each response. The qualitative findings from the FGDs to examine students' opinions towards MT, MTPE and HT will be presented in 6 sections that correspond to the sub-research questions. Responses of some interviewees will be given under each question from the FGDs as examples of the identified themes. Each question will be written in bold followed by the responses of the students. The FGDs were conducted in Arabic and translations of the students' responses are my own. A summary of the findings is provided at the end of this section.

#### 4.2.1.1 MT Use: How often do you use MT?

The aim of this question is two-fold: firstly, it would provide a background about students' experience with MT. Secondly, the responses to this question would be used to modify the contents of the teaching intervention. The degree of MT experience that the students reveal would inform the time allocated for an introduction about MT and MT use at the beginning of the

teaching intervention. Such gathered detail would save time and allow more time for focused PE skills in case the students showed moderate to high experience with MT.

At the beginning of each FGD session, students were asked about their MT/PE experience. The thematic analysis of the significant responses regarding MT use identified five themes that characterise the various uses of MT. The identified themes are summarised in Table 4-2 below.

Table 4-2 Identified themes of MT use

| Code | Theme |
|---|---|
| MT use | Always use MT |
| | Sometimes use MT |
| | Use HT/ do not use MT |
| | Use MT to meet tight deadlines |
| | Use MT for specific purposes |

Responses from students who indicated that they 'always used MT' (6 students) showed that the main reason for using MT was tight deadlines. Other responses indicated that MT helped the students with the guessing of the terms within a specific context. A student who said that she always used MT shared her experience about people warning her against using it as it might impact negatively upon her HT skills:

> (P3) I use Google Translate all the time in everything, literally in everything related to translation courses. I also used it in my 'graduation project' training period. Of course, with time you gain the experience to identify MT errors. Thankfully, using it did not affect me as many have told me 'do not use Google Translate. If you get used to it, you will not be able to translate by yourself in translation exams', but it did not affect me.

About two thirds of the students (17 students) indicated that they 'sometimes' used MT. Their additional comments, however, showed that they often considered post-editing to be 'working with a translation tool', including editing translation memory (TM) matches as well as MT output. Therefore, the opinions on post-editing taken from the focus group discussions may encompass issues related to the usage of translation tools in general, in addition to post-editing. Some example answers to the open question 'how often do you personally use MTPE as a translation method' can be seen below.

> (P1) I sometimes use Google Translate but I never rely on it. I prefer my own translation because when I see MT output, I fail to edit it, and I tend to take it as it is.
>
> (P13) Honestly, I prefer human translation. I use electronic dictionaries to look up unknown words. But to be honest, if there is not enough time, I use MT for gisting so that I understand the whole text before I translate the whole text by myself.

Three responses which revealed a tendency not to use MT showed three different patterns: students did not use MT because they preferred HT, students used MT for specific purposes only while preferring HT, and students used MT only to meet tight deadlines.

In addition to this question serving as an introduction to the discussions, the responses of students showed that the majority were familiar with MT and had some experience using it (6 students indicated that they always used MT while 17 were familiar with it and sometimes used it). This provided a clear indicator that the teaching intervention would need to provide only a very brief introduction about MT before focusing on post-editing skills.

#### 4.2.1.2 Degree of Reward: How rewarding is PE compared to HT?

To define the scope of the term 'reward', students were informed that *a translation method is considered 'rewarding' when the translator acknowledges its value(s) in their translation tasks.* Students shared their opinions about how rewarding they thought PE was compared to HT. The thematic analysis of the significant responses regarding the degree of reward identified nine themes that characterise the various opinions shared by the students. The identified themes are summarised in Table 4-3 below.

Table 4-3 Identified themes in the degree of reward

| Code | Theme |
|---|---|
| Degree of reward | HT is more rewarding- translation style |
| | HT is more rewarding- MT use only for gisting |
| | HT is more rewarding- MT output is full of errors |
| | HT is more rewarding- MT decreases job opportunities |
| | HT is more rewarding- HT is faster |
| | MTPE is more rewarding- faster than HT |
| | MTPE is more rewarding- financially more rewarding |
| | MTPE is more rewarding- it can be post-edited |
| | MTPE is more rewarding- MT will be powerful |

Students' opinions were split between HT and MTPE where 11 participants thought HT was more rewarding while 12 participants thought MTPE was more rewarding. Three participants did not share their opinions in this part of the discussion. As the ethical approval obtained for this part of the study did not include student observation, I could not tell why these three students refrained from sharing their opinions.

Those who thought HT more rewarding were mainly students who 'did not' use MT/PE or 'sometimes' used it when asked about 'MT use'. They supported their opinions about the degree of reward with several different reasons. One of the students thought that one of the most important values in translation is the style and clarified that this is why she preferred HT:

> (P10) I don't like MT. It is more rewarding to translate by myself because post-edited MT will not show my style of translation.

Another reason why students thought HT was more rewarding had to do with the students' level of reliance on MT output. Two examples show how little some of the students rely on MT output: They declared that they would only use it for gisting, or in another case, a student preferred seeking assistance from another human because she did not trust MT output:

> (P13) HT is the most rewarding method. I only use MT for gisting.
>
> (P14) Translating by myself is more rewarding. Sometimes when a sentence is not clear, I seek the help of my classmates, but I do not refer to MT. It is simply not reliable.

Students who answered 'often use MT/MTPE' and still preferred HT provided explanations for why they did not prefer MTPE as they thought it contradicts the values of financial reward and accuracy. They thought that MT risks job offers, and that MT output is full of errors.

> (P11) It is true that it [MT] may decrease job offers for translation graduates as they say.
>
> (P23) I don't like to be negative. I do not think MT would overtake our roles as translators, but it would seriously reduce our chances in the job market in the future. I am pro-technology and advancement in the field even if they lessen our chances. It does not matter; technology is more important.
>
> (P15) I like Google Translate although it has numerous errors. HT is professionally more rewarding because MT does not understand.

One student who thought HT is more rewarding supported her opinion by revealing that although MT is usually preferred for its fast outcome, that she actually thought that HT was faster.

> (p2) I remember I once used MT and tried to post-edit the output. It took me so much time: I spent time trying to edit the text then when I decided to manually translate the original text, because the quality was way too poor, I was stuck with the MT output that I had received, and it was so hard for me to forget it and get a fresh start! I would definitely finish more translation jobs through HT alone.

However, some of those who thought MTPE is more rewarding believed so because they thought that MTPE is faster than HT (6 students) that it is more financially rewarding because clients do not care much about the final quality.

> (P4) I read that many clients do not search for a high-quality translation. They just want someone who is cheaper and faster to translate into Arabic. I anticipate that MTPE is more rewarding because people do not care about quality. They just want fast services.

One student had a more positive view of MTPE. Particularly, she was more positive about the fact that translators can actually edit MT output. She affirmed that MTPE is more rewarding because the output could be 'dealt with' or edited. Another student had an extreme thought about why she believed MTPE was more rewarding as she thought that MT will reach a point where no human effort is required for the improvement of its output. She commented:

> (P26) Yes, MT will improve. I believe the 'robot' will improve autonomously without the humane interference. It will improve to the extent of exceeding human intelligence or at least competing against it.

In conclusion, although this question was about the degree of reward of the different translation methods, a number of themes emerged when examining the students' responses: 8 students mentioned 'MT errors' as the reason why they thought HT was more rewarding. 7 students talked about their fear of 'MT jeopardising job offers'. Whereas 6 students thought MT was more rewarding because it mainly 'saves time' while only one student thought that MTPE takes more time than HT. Four students thought MT is more rewarding because of its tendency to be post-edited resulting in high quality texts. Those who preferred HT mentioned creativity and freedom as their important factors to believe HT was more rewarding.

Students in the FGDs have engaged in the conversation about the degree of reward and expressed verbally their opinions about why they thought HT is more rewarding than MTPE or vice versa. All in all, the discussion about the theme ''degree of reward'' showed that students have mixed opinions (11 students for HT and 12 students for MTPE) but when reflecting on these

students' answers to the previous question about their MT experience, I noticed that those who 'sometimes' and 'often' used MT were the ones who recognized the benefits of implementing MTPE in the daily life of the translator. Further discussion about these findings is presented in section 5.2.1 below.

#### 4.2.1.3 Usefulness: How useful is MT output according to translators?

In this section, participants shared their opinions about the usefulness of MT output. The thematic analysis of the significant responses regarding the usefulness of MT output identified five themes that characterise the various opinions shared by the students. The identified themes are summarised in Table 4-4 below.

Table 4-4 Identified themes regarding usefulness

| Code | Theme |
|---|---|
| Usefulness | MT is useful because it is fast |
| | MT is useful because it provides field-specific terminology |
| | MT is less useful because it misses the skopos of the text |
| | MT usefulness relies on the purpose of the translation |
| | MT is useful with limited text types |

Again, students' opinions were split between those who thought MT output was less useful (14 students) and those who thought it was often useful (12 students).

Two out of the twelve students who thought that MT output is often useful explained that this is true when MT is followed by post-editing because MT output provides field-specific terminology and because MTPE is faster than HT. Their added clarifications can be seen in the examples below:

> (P15) I believe that if MT is followed by post-editing, the MT output is more useful because it can provide me with the most frequently used terms and phrases within the specific field.

> (P8) MT output is useful when combined with PE because it really is fast while HT is all about creativity. It is impossible to get creative when relying on the quality of MT output alone, but MT quality is acceptable. I honestly rely on its usefulness.

On the other hand, a student thought that MT output is less useful because it misses what Vermeer (1989) would refer to as the Skopos of the text in a way that makes the translator's job even more difficult when editing:

> (P12) MT output can be deceiving because sometimes it removes or changes the purpose of the text and the translator either does not pay attention to the

> skopos or finds post-editing the translated text even harder than translating it from scratch.

A student made it clear that the usefulness of the MT output depends on the purpose of the translation but that there are limitations to its usefulness:

> (P6) MT output does the job when you are gisting the meaning but not more than that. I believe MT output can be useful or useless depending on the purpose of the translation.

Another student added that this is especially true in the case where MT is used in certain organizations and or companies where the text types are fixed:

> (P25) If you work in an organization or a ministry that has a specific format of texts like law firms or petroleum companies, MT out can be extremely useful because you are actually sharing a unified database that everyone has agreed upon using.

A student believed that the usefulness of MT has largely to do with the need for a fast turnaround when a translator post-edits MT output for large corporations:

> (P7) I think MT output is more useful for the translator who works in a large corporation on large projects where they need to work fast. MT would save 80% of the time and all that is left for me is to post-edit.

However, those students who thought that MT output is less useful related this to different reasons. Some thought that dealing with MT, i.e., using MT output would badly affect their translation skills:

> (P2) I don't use MT so that I do not lose my linguistic competence.

Not only the use of MT would jeopardize the translators' linguistic competence according to those who thought MT output was less useful, but it is also insensitive to cultural differences between source and target languages:

> (P11) My brain can synthesize the translation into a target culture. MT output is useless in this regard. MT does not compare to human translation.

All in all, opinions regarding the usefulness of MTPE were diverse and students supported their opinions with examples from real work environments that involved MT use. Discussion of the findings is provided in 5.2.1.2.

#### 4.2.1.4 Speed: Which translation method is perceived as being faster?

This category focused on exploring which translation method the students perceived as being faster and why. The thematic analysis of the significant responses regarding speed identified six themes that characterise the various opinions shared by the students. The themes are summarised in Table 4-5 below.

Table 4-5 Identified themes regarding speed

| Code | Theme |
|---|---|
| Speed | MTPE is faster because MT does most of the translation work leaving little for PE |
| | HT is slower because of the cognitive effort associated with it |
| | MTPE is slower because it requires cognitive effort in complex texts |
| | MTPE is faster when the translator masters PE skills |
| | Speed depends on the text type |
| | MTPE is faster in complex texts because it provides gisting |

In total, 25 students responded to this question. 18 students thought that MTPE is generally faster while 4 students believed that HT is as fast as MTPE. 3 students thought HT is faster than MTPE. Their added justifications can be found below.

Some students expected PE to be generally faster because the percentage of text that is left for editing is much shorter than when translating from scratch:

> (P23) I expect PE to be faster because you only have to post-edit 30-40% of the text. Thus, you will focus on a third of the text which requires much less time than translating the whole text from scratch.

Others suggested that HT requires more cognitive effort:

> (P26) In my opinion, MTPE is faster for many reasons but most importantly because sometimes, when a person is manually working on translating a text, they tend to overthink the meanings which might not be a required skill for that specific text. When I use MT, it provides me with the general meaning from the first moment, which saves time.

On the contrary, a student thought that MTPE required more cognitive effort and that it is easier to handle the translation when manually translating complex texts:

> (P10) When the text is 'difficult', MT would generate an output that requires much more time to fix. In this case, translating from scratch is easier for me. I believe speed is relative and that it depends on the complexity of the text.

According to a student who thought MTPE is faster when the translator masters the skills required to post-edit, it is much faster for them to spot the parts in the text that require editing. Thus, their time is saved, and their productivity is increased:

> (P14) I view MTPE as the faster method because once the translator becomes an experienced post-editor, the whole process of translating a text is faster. PE is just a skill that once mastered, faults and mistakes will be spotted quickly, and the translation task will be completed faster giving the translator a chance to work on more tasks per day.

Also, for some students, text type is a major reason why MTPE is faster because when a translator is dealing with a text type of a technical nature, it is much faster to post-edit it than to translate from scratch every time:

> (P19) In my opinion, MTPE is faster because it saves me from searching for new terms and vocabulary items. It also provides acceptable translations of full sentences especially when dealing with texts of a technical nature such as legal translation. This way, PE is faster as it provides me with more options and all I'm supposed to do is to give the text a final look to ensure it is well-edited.

In addition to technical texts, it is also thought that when a translator is handling an unfamiliar text type, MTPE is faster because it provides a general overview of the translation which helps the translator grasp the overall meaning without wasting time on searching in dictionaries:

> (P11) Of course, when the text is unfamiliar, MT would help me get the translation job done faster because it would provide me with a "translated text'' which I will edit through looking for the translation faults here and there [pause] this word might be changed, and that phrase has to be corrected and so forth.

There seems to be some correlation between finding MTPE more rewarding and believing it is also faster, with none of the students who thought MTPE was more rewarding thinking that it is slower than HT. However, eight out of the nine students who thought that HT is more rewarding thought that MTPE is faster than HT which indicates no correlation between HT degree of reward

and how fast it is in students' opinions. Nonetheless, a student expected longer texts to be translated faster through HT because according to her:

> (P25) HT is faster because the time you are using depends on the word count of the text. If you use MT to translate a long text, then work on post-editing the output, you will have to re-read the text and the output anyway then post-edit, so you are basically wasting time. While if you translate the same text manually, you will start to translate immediately without wasting time, and you will achieve faster results and all you need is to give it a second final look at the end.

#### 4.2.1.5 Quality: How is the final quality of both methods of translation perceived?

In addition to speed, this study is interested in exploring the perceived quality of the post-edited texts. The thematic analysis of students' responses has identified three main themes that characterize the responses from the students. The identified themes are summarized in Table 4-6 below.

Table 4-6 Identified themes regarding quality

| Code | Theme |
|---|---|
| Quality | MTPE does not deliver the ST message properly |
| | MTPE is resourceful |
| | MT output is full of errors |

Most of the students in the focus group discussions did not seem convinced of the quality of texts resulting from MTPE in comparison to that of HT texts. 15 students thought that HT generates better quality, while 7 students thought MTPE has a better-quality outcome. Four students thought that the quality of both methods is comparable.

The majority of students believed HT quality is better because MTPE adopts a semantic approach which does not effectively convey the ST message and that a human translator can adopt a freer approach that will convey it better:

> (P1) In the last assignment we were given, some students post-edited MT output. Their translations were awful. They were word-for-word. They did not deliver the message of the author. On the other hand, the part that I translated was hard to criticize because I manually translated the text. Thus, I, of course, prefer HT quality.

> (P3) The quality of HT is better even if the quality of MTPE is good. If you give five translators a text and they all post-edited the MT output, the outcome will be five translations of similar style. The translator's touch will be absent. In HT, the quality is surely better, and the translation shows the style of a human not a machine.

> (P5) When I go to buy a translated book, I ask for certain names because I like their translation style. This is why HT is better in term of quality.
>
> (P9) Generally speaking, I feel that HT generates better quality. If I use MT, I must edit the output, but I don't depend on this method. I like HT because it adds beauty to the translation. I feel that this makes a difference even with the client.
>
> (P7) No matter what I do when post-editing MT output, the translation is usually literal!

On the other hand, some students though that the quality of MTPE is better because it is resourceful:

> (P10) In my opinion, MTPE quality is better because the work would be the result of two brains, i.e., me and the artificial intelligence. This way, I have referred to two resources in order to come up with the best possible quality.
>
> (P12) MTPE quality is better because it provides more resources so that I would know what specific terms are used in a specific field and how things are expressed in that particular field.

However, for those who believed that the quality of HT is better, it was due to the many errors found in MT output and the tendency of overlooking them by translators:

> (P4) Sometimes, you would doubt prepositions that are machine generated because machine translation does not consider collocation.
>
> (P6) Sometimes you see your errors when you are translating manually. When using the MTPE method, there are usually errors that affect the overall quality which you as a translator cannot spot.

In regard to translation quality, although some students gave MTPE the credit of being resourceful and useful as a reference for second opinion, the majority have actually criticized the inability of MTPE to deliver the ST message properly. Some students raised the issue of MT output being full of errors and that such errors are usually missed when post-editing. Further discussion of the findings is provided in 5.2.1.5.

#### 4.2.1.6 Preference: Which translation method is the most preferred translation method?

Towards the end of the focus group discussions, I asked the students about their preferred translation method (HT or MTPE). Three major themes were identified through the thematic analysis of the responses. These themes are summarized in Table 4-7 below.

Table 4-7 Identified themes regarding preference

| Code | Theme |
|---|---|
| Preference | MTPE causes feelings of guilt |
| | MTPE helps meet tight deadlines |
| | Preference depends on the complexity of the text and the translator's familiarity with the field |

Generally speaking, most students preferred HT over MTPE. Out of 26 students, 19 preferred HT while 7 preferred MTPE. Nine of those who thought HT is more rewarding also preferred using it. However, one student who thought HT is more rewarding preferred MTPE as a translation method.

Students' discussions focused on three concepts that determined their preferences: ethical considerations, translation productivity, and the complexity of the text. Those who spoke about how MTPE made them feel uncomfortable explained that they usually felt guilty when they post-edited MT output:

> (P13) Honestly speaking, I prefer HT because it makes me feel good about myself and my skills. When the timeframe for delivering the translation is too tight, I post-edit MT but I feel guilty and awful!
>
> (P24) I am soon to be a fresh graduate and I have got great respect for my work ethics. Maybe when I am more experienced and receive loads of texts to translate, I might overlook those ethics and refer to MTPE.

Others preferred MTPE for when they are under pressure and need to deliver the translation fast:

> (P16) I prefer HT but when I am under pressure, which depends on how much free time I have got, I tend to post-edit MT output because it is fast. It saves both my time and effort.
>
> (P17) I prefer to use MTPE to complete home assignments during exam times in order to be able to spare time for studying.

Finally, for those who preferred a certain translation method depending on text types, it mainly is dependent on the complexity of the text and the translator's familiarity with the field:

> (P24) My preference depends on the complexity of the text. Some texts are super complex that I cannot understand them, and MT provides me with a general idea to help me understand the meaning [pauses] yes, my preference depends on the complexity of the text.

> (P27) If I am familiar with the field, I prefer to manually translate because I enjoy translation more. Those fields that I am forced to translate into, I usually prefer MTPE to complete the task.

#### 4.2.1.7 Summary of Findings from FGDs

All in all, the analysis of the FGDs responses revealed that students held mixed opinions about HT, MT and PE. In terms of the degree of reward, students were split between HT and MTPE. Those who thought MTPE was more rewarding listed the following reasons to back up their opinions: MTPE is faster than HT, clients prefer productivity over quality, MT output can be edited, and MT output is constantly improving. On the other hand, those who thought HT is more rewarding thought so because: HT quality is better, MT output is full of errors, HT brings about better translation style, and one student thought that HT is actually faster than MTPE.

As for how useful students thought the two translation methods were, students thought that MTPE is useful for gisting, to save time, to increase productivity and to provide alternative TTs for comparison. However, those who thought MTPE was not useful, said they thought it would affect their HT skills negatively, that it was literal, and that the context can easily be lost.

When asked about their thoughts on the speed of the two translation methods, students generally thought that MTPE was faster, nevertheless, some students provided more in-depth details about: the text type (if text type is not familiar, MTPE makes translation faster through gisting), text complexity (more complex texts require more temporal effort and vice versa), and finally, the concept that speed is proportional, i.e., MT output is instant, but PE may require more time. Overall, students thought that the longer the texts, the more the required temporal effort to post-edit them.

The question about translation quality of both methods showed more negative opinions about MTPE. Students supported their opinions with the following list of explanations about MTPE: it is literal, it jeopardises the translator's style, it is full of errors, the over-reliance of translators on MT affects their ability to notice errors, it damages the cohesion of Arabic texts, and it may suggest wrong meanings risking the overall accuracy of TTs.

Finally, when students were asked about their personal preferences, most of the students revealed that they did not prefer MTPE for the following reasons: MT is rather a gisting tool, it may damage HT skills, it might decrease confidence in HT skills, it does not take cultural differences into account, it is dishonest, and that it is against work ethics. Discussion of these findings as well as those resulting from the survey are discussed in section 5.2.1.

### 4.2.2 Results from Retrospective Pre-test Survey (Post)

A comparison of pre-and post-intervention of opinions within the experimental group was conducted through the use of a retrospective pre-test survey. The seven questions about opinions in the survey were analysed using percentages. Since the answers are defined based on an ordinal scale, the Wilcoxon test (non-parametric approach) for two paired samples was used to test whether the opinions of the students after the teaching intervention significantly changed from before. The test was applied to each question (See Table 4-8 below).

Table 4-8 below examines pre- and post-survey of change in opinion within the experimental group (27 students). As previously explained in 3.10.2.1, calculating the Z-scores provides an interpretation of the effect size (if the Z-score is positive, it indicates positive change between pre and post responses and vice versa). The statistical analysis of the survey responses shows that the mean post-survey Z-scores of six statements for the group were all greater than 2.0. This means that students' opinions about MTPE have changed positively. The one statement that showed no significant change was the one about speed which indicates that students' opinion that MTPE is faster than HT remained the same after the intervention. Possible interpretations are provided in the discussion 5.2.1.

Table 4-8 Summary of students' results

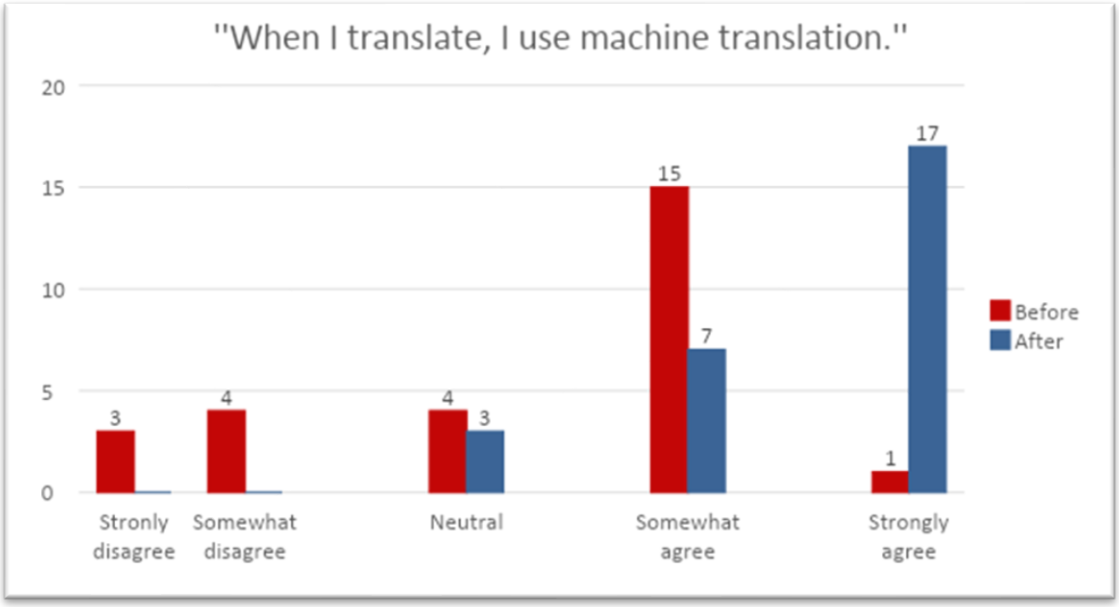
| Question | | Pre | | Post | | Wilcoxon test | |
|---|---|---|---|---|---|---|---|
| | | N | % | N | % | Z | p-value |
| 1.When I translate, I use machine translation | Strongly Disagree 1 | 3 | 11.1% | 0 | 0.0% | 4.048 | <.001 |
| | Somewhat Disagree 2 | 4 | 14.8% | 0 | 0.0% | | |
| | Neutral 3 | 4 | 14.8% | 3 | 11.1% | | |
| | Somewhat Agree 4 | 15 | 55.6% | 7 | 25.9% | | |
| | Strongly Agree 5 | 1 | 3.7% | 17 | 63.0% | | |
| 2.'I find post-editing (------) than manual translation | more rewarding 1 | 2 | 7.4% | 21 | 77.8% | 4.304 | <.001 |
| | as rewarding 2 | 9 | 33.3% | 5 | 18.5% | | |
| | less rewarding 3 | 16 | 59.3% | 1 | 3.7% | | |
| 3.I find that the output of machine translation systems is (--------). | Useful 1 | 5 | 18.5% | 22 | 81.5% | 3.564 | <.001 |
| | neither useful nor not 2 not useful 3 | 8 | 29.6% | 2 | 7.4% | | |
| | not useful | 14 | 51.9% | 3 | 11.1% | | |
| 4.I find that the output of machine translation systems is (--------). | easy to handle 1 | 6 | 22.2% | 15 | 55.6% | 2.355 | .019 |
| | neither easy nor difficult 2 | 14 | 51.9% | 8 | 29.6% | | |
| | difficult to handle 3 | 7 | 25.9% | 4 | 14.8% | | |
| 5.I think I translate (------) when translating manually than when post-editing. | Faster 1 | 8 | 29.6% | 12 | 44.4% | .038 | .969 |
| | as fast 2 | 9 | 33.3% | 2 | 7.4% | | |
| | Slower 3 | 10 | 37.0% | 13 | 48.1% | | |
| 6.I think the quality of my translations is (------) when post-editing than when translating manually | Worse 1 | 14 | 51.9% | 2 | 7.4% | 4.053 | <.001 |
| | Similar 2 | 10 | 37.0% | 6 | 22.2% | | |
| | Better 3 | 3 | 11.1% | 19 | 70.4% | | |
| 7.I prefer | PE over HT 1 | 4 | 14.8% | 22 | 81.5% | 4.243 | <.001 |
| | HT over PE 2 | 23 | 85.2% | 5 | 18.5% | | |

#### 4.2.2.1 MT Use

Twenty-seven students were asked about the frequency of their MT use both before and after the teaching intervention: 'When I translate, I use machine translation'. After the intervention, the distribution of responses ranged only between ''neutral'', ''somewhat agree'' and ''strongly agree'' (no student chose ''strongly disagree'' or ''somewhat disagree'') (Table 4-8, Figure 4-1). ''Neutral'' responses about MT use decreased from m=4 (14.8%) before the intervention to 3 (11.1%) after the intervention, and responses of ''somewhat agree'' to use MT sharply decreased from n=15 (55.6%, before) to n=7 students (25.9%, after) in favour of ''strongly agree'' to use MT raising from n=1 (3.7%, before) to n=17 students (63%, after).

The latter percentage suggests that the intervention was successful in demonstrating the integration of MT use in the daily work of translation students. Using the Wilcoxon test, there is a significant positive change in students' opinions after intervention (Z=4.048, p-value<.001).

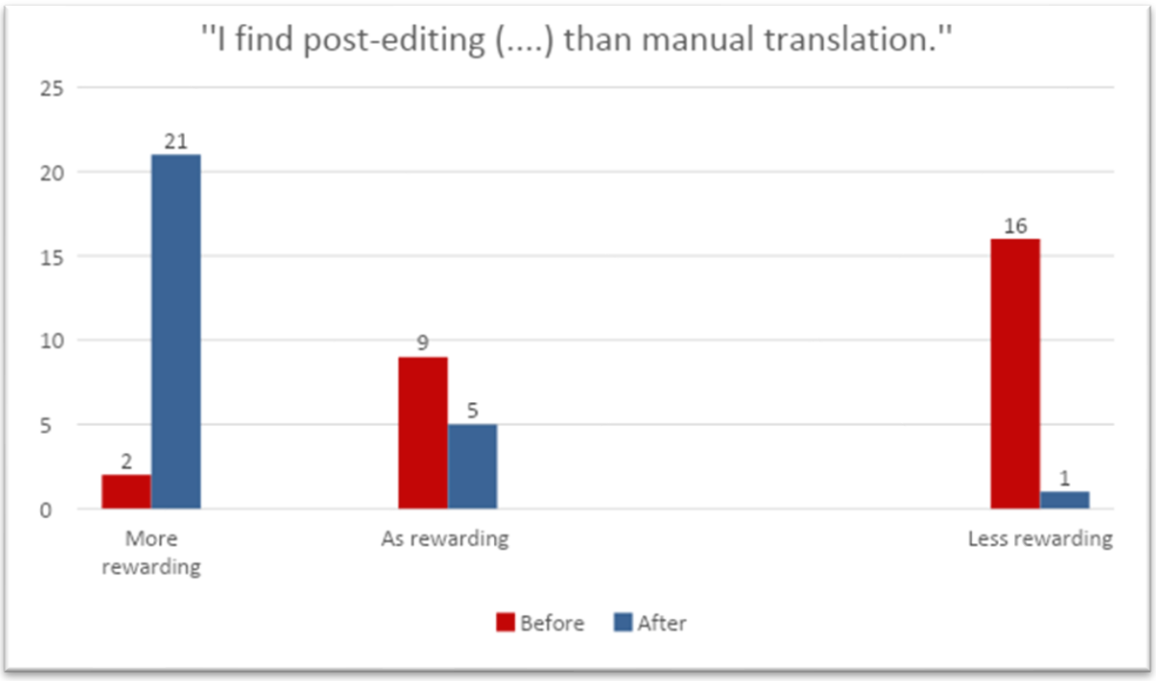
Figure 4-1 MT use




#### 4.2.2.2 Degree of Reward

Students were asked to fill in the statement 'I find post-editing (------) than manual translation' with one of three choices: ''more rewarding'', ''as rewarding as'', or ''less rewarding''. After the intervention, the distribution of responses has almost changed to the exact opposite with 21 students (77.8%) thinking that PE is ''more rewarding'' than HT while one student (3.6%) thinking that PE is ''less rewarding'' than HT (Table 4-8, Figure 4-2). The number of students who thought that both methods of translation are ''as rewarding'' dropped from 9 (33.3%) to 5 students (18.5%) which again indicates that the intervention was successful in demonstrating the degree of reward of PE in the daily life of a translation student. Using the Wilcoxon test, there is a significant positive change in the attitudes after the intervention (Z=4.304, p-value<.001).

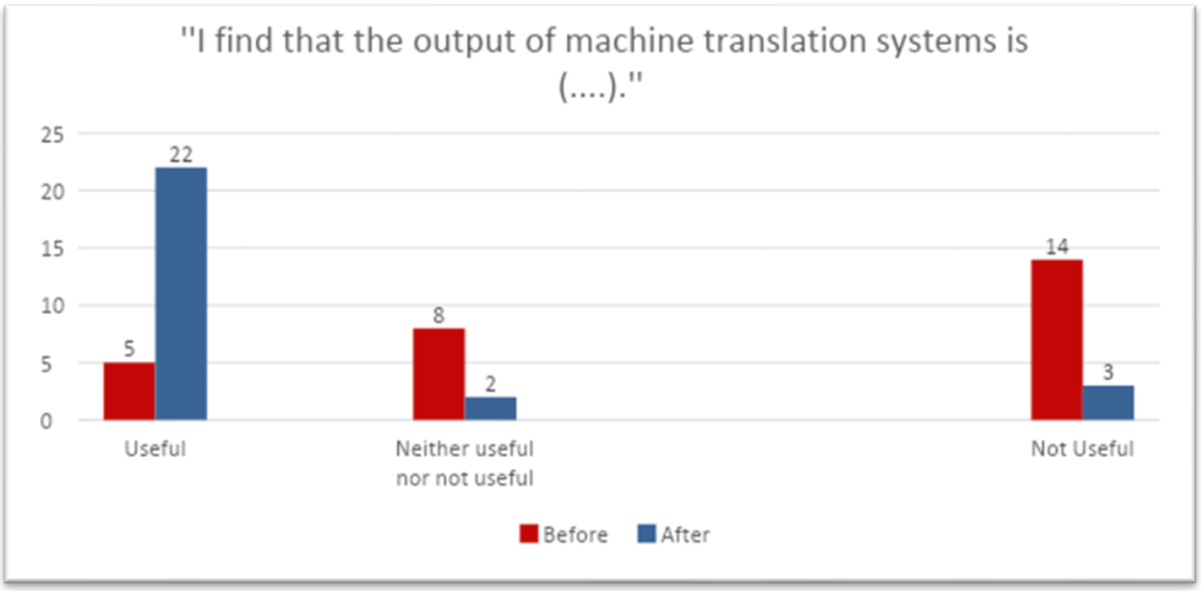
Figure 4-2 Degree of reward




#### 4.2.2.3 Usefulness

Usefulness of MT output is important in the process of post-editing as the adoption of new technology depends on it (Davis 1989). So, students were asked to choose from ''useful'', ''neither useful nor not useful'', or ''not useful'' in response to 'I find that the output of machine translation systems is (--------)'. After the teaching intervention, 22 students (81.5%) thought that MT output was useful while the number of those who thought it was ''neither useful nor not useful'' dropped from 8 to 2 students (29.6% to 7.4%) (Table 4-8, Figure 4-3). The number of students who thought MT output was ''not useful'' sharply decreased from n=14 (51.9%) students to 3 students (11.1%) (Table 4-8, Figure 4-3). Using the Wilcoxon test, there is a significant positive change in the attitudes after intervention (Z=3.564, p-value<.001).

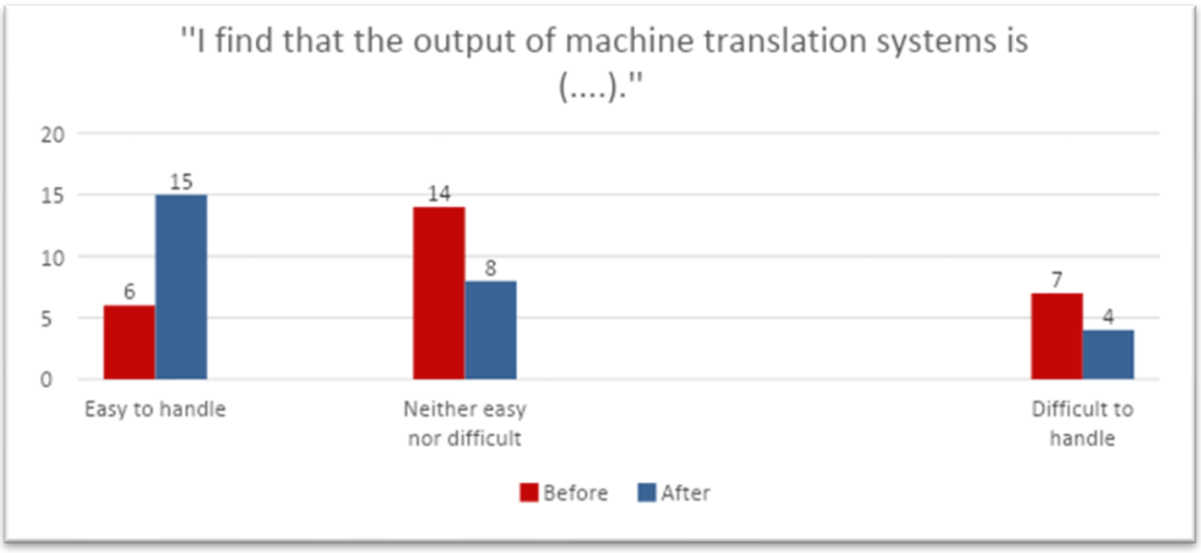
Figure 4-3 Usefulness




#### 4.2.2.4 Ease of Use

For translators to be able to process post-editing successfully, they ought to find the MT output easy to handle (Daems 2016). When students were given the statement 'I find that the output of machine translation systems is (--------)', before the intervention, half of them (n= 14, or 51.9%) thought that MT output was ''neither easy nor difficult'' to handle. 7 students who thought that MT output was ''difficult to handle'' (28.6%) were slightly higher than those 6 students who thought it was ''easy to handle'' (22.2%) (Table 4-8, Figure 4-4). Opinions of the students after the intervention indicate that the teaching intervention succeeded in promoting methods of handling the MT output as the number of students thinking that MT output was ''easy to handle'' increased to 15 (55.6%). Almost one third of students (n= 8, or 29.6%), however, thought that it was ''neither easy nor difficult'' while 4 students (14.8%) thought that it was ''difficult to handle''. Using the Wilcoxon test, there is a significant positive change in students' opinions after the teaching intervention (Z=2.355, p-value=.019)

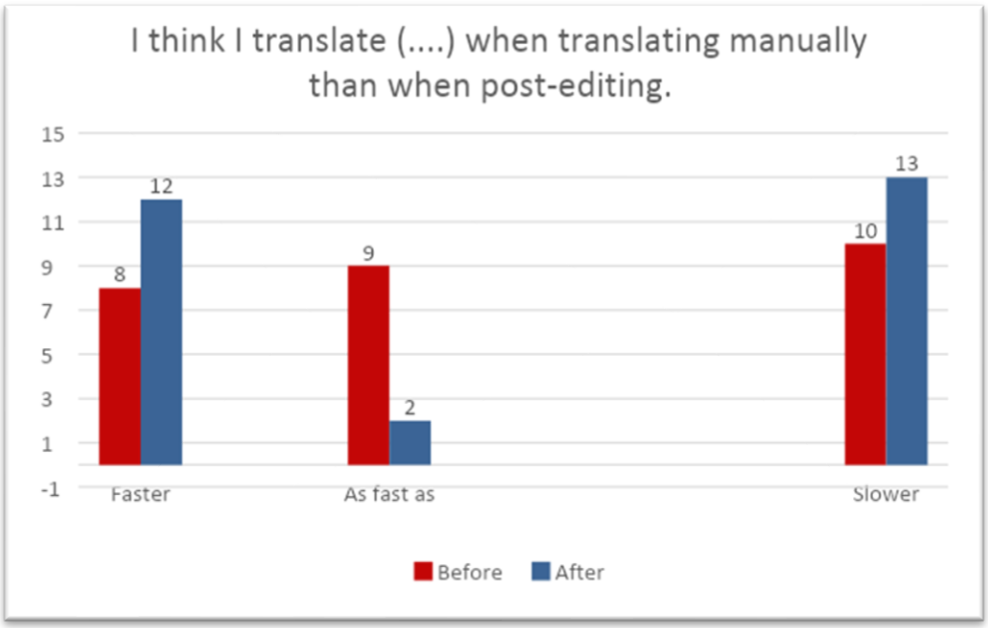
Figure 4-4 Ease of use




#### 4.2.2.5 Speed

When it comes to the perceived speed of each translation method, opinions of students who took part in the MTPE teaching intervention were a bit more diverse. When students were given the statement 'I think I translate (------) when translating manually than when post-editing'. Before the experiment, 8 students (29.6%) thought they translated ''faster'' when translating manually. A slightly higher percentage of students (n= 9, or 33.3%) thought they translated ''as fast'' in both translation methods while a slightly higher percentage of students (n= 10, or 37%) thought that they translated ''slower'' when working manually than when post-editing MT output. After the students were given the chances to put both translation methods into practice during the teaching intervention, their opinions seemed to take sides rather than being neutral as 12 students (44.4%) of students thought they translated ''faster'' when manually translating a text. A slightly higher percentage thought otherwise as 13 students (48.1%) thought that translating manually was ''slower''. Two students (7.4%) thought that both translation methods were ''as fast'' (Table 4.8, Figure 4-5). Using the Wilcoxon test, there was not any significant change in students' opinions after the teaching intervention (Z=-.038, p-value=.969).

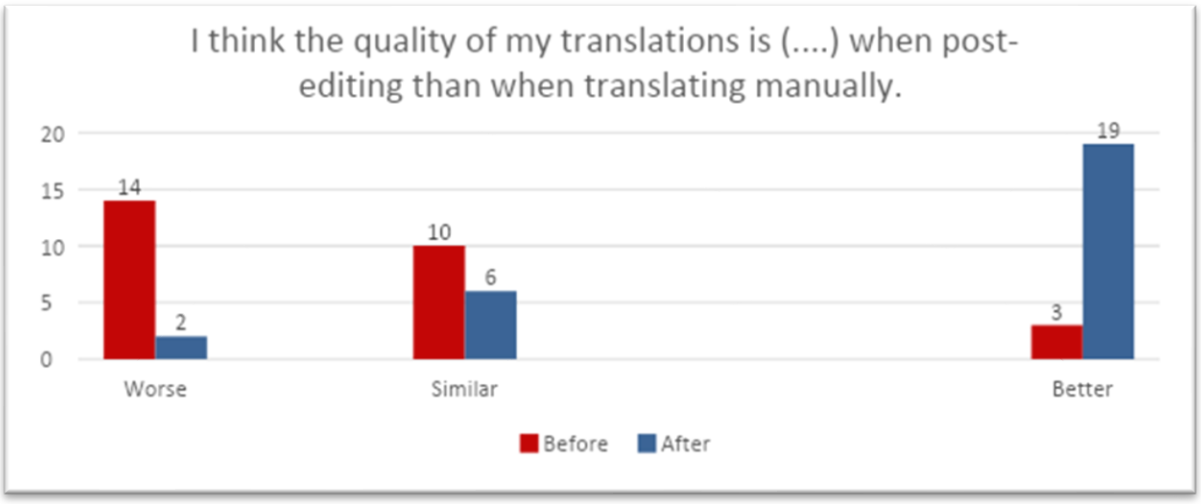
Figure 4-5 Speed




#### 4.2.2.6 Quality

During the teaching intervention, students were presented with common MT error patterns in the Arabic language as well as the different styles of texts that can be post-edited. When asked about their perceived quality of translation 'I think the quality of my translations is (------) when post-editing than when translating manually', before the intervention, half of the students (n= 14, or 51.9%) thought that the quality of their translations was ''worse'' when they post-edited while 10 students (37%) thought the quality was ''similar'' to that when they translated manually. Three students (11.1%) thought they managed the quality well and that they produced ''better'' quality when post-editing MT output. After the intervention, the number of students thinking that they managed ''better'' quality through post-editing MT output raised from 3 to 19 (11.1% to 70.4%). Both responses about PE quality being ''worse'' or ''similar'' have decreased respectively from 14 to 2 (51.9% to 7.4%) for ''worse'' quality, and from 10 to 6 (37% to 22.2%) for ''similar'' quality (Table 4-8, Figure 4-6). Using the Wilcoxon test, there was a significant positive change in students' opinions after the intervention (Z=-4.053, p-value<.001).

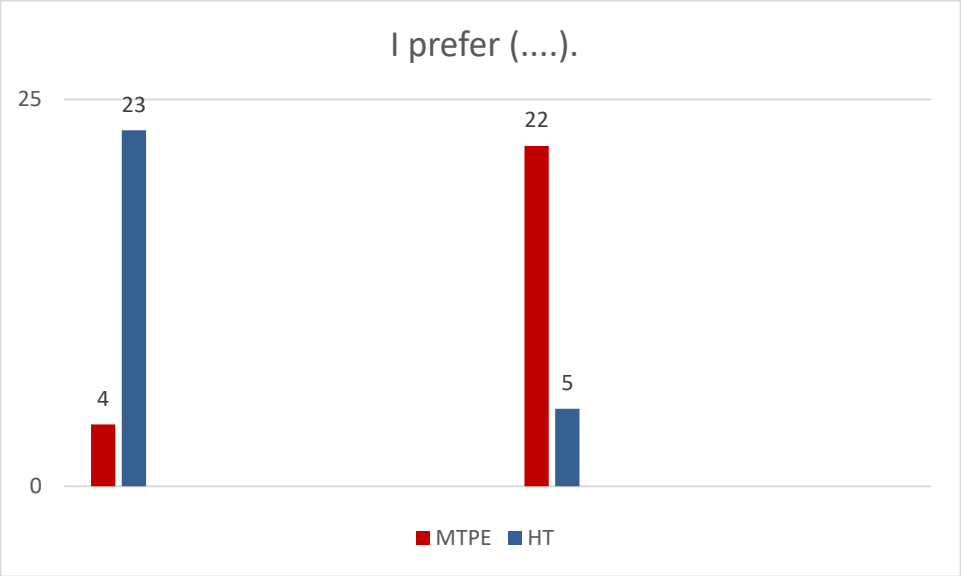
Figure 4-6 Quality




#### 4.2.2.7 Preference

Since personal preference of translation method is considered an important factor in determining the success of the process (Daems 2016), it was worth asking the students about their preferred method of translation. When asked to fill in 'I prefer -----' with two choices to choose from 'PE over HT'' or ''HT over PE'', before the teaching intervention, 23 students (85.2%) preferred ''HT over PE'' while 4 students (14.8%) preferred ''PE over HT''. However, after the teaching intervention, the percentage of responses preferring ''HT over PE'' decreased to (n= 5, or 18.5%) while the percentage of responses preferring ''PE over HT'' raised to 22 (81.5%) (Table 4-8, Figure 4-7). Using the Wilcoxon test, there was a significant positive change in students' opinions after the teaching intervention (Z=-4.243, p-value<.001).

Figure 4-7 Preferred method of translation

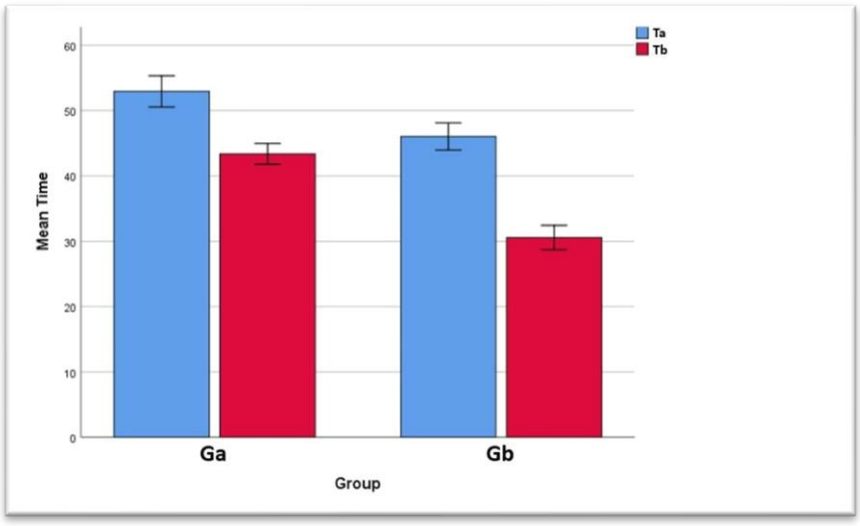



In summary, the results from the survey that was conducted after the intervention revealed that the students in general showed a shift towards significantly more positive opinions about MTPE training and use except for opinions about speed which generally remained the same. A further discussion of the survey results is provided in 5.2.1.

## 4.3 Translation Productivity Results

This section reports the results of translation students regarding their productivity in translation tasks. The productivity was measured through comparing results drawn from the control group (Ga) and the experimental group (Gb) in the pre-test (Ta) and the post-test (Tb). As previously mentioned, productivity is calculated through a comparison of the total task duration in minutes to complete the translation task for both Ta and Tb. Paired sample *t*-tests as well as repeated measures ANOVA were run to compare the productivity of HT and MTPE in both Ga and Gb in different intervals of time (i.e., Ta and Tb). The quantitative results will be presented in three sections: The first section describes results from the longitudinal comparison of Ga (paired sample *t*-test), the second section describes results from the longitudinal comparison of Gb (paired-sample *t*-test), and the third section describes results from the horizontal comparison of both Ga and Gb in Ta and Tb (repeated measures ANOVA) in an attempt to answer RQ2 ‘’What are the differences in the process between HT and MTPE? Is the productivity gain from post-editing MT output larger than that of HT?’’

Figure 4-8 Total translation time (Ta vs. Tb) for Ga and Gb




### 4.3.1 Productivity Gain in Ga

In this section, the productivity gain in Ga was compared to provide a reference for comparison with the productivity results from Gb. Logically, students are expected to gain more productivity in HT after 4 more weeks of training. Since data was normally distributed, the paired-samples *t*-test was conducted on students in Ga to estimate their productivity in HT at one point of time (Ta) and again after 4-weeks (Tb). The mean decrease in total translation time change scores was 10 minutes (Figure 4-8 above, Table 4-9 below). There was a statistically significant decrease in translation pre-test total time spent (Mean= 53 minutes, SD= 12.87) to post-test total time spent (Mean= 43 minutes, SD= 8.54), t (29) = 3.76, p=.001). Therefore, the results from Ga show that there was a productivity gain of 19% in Tb when compared to Ta.

Table 4-9 Productivity gain in Ga (Ta and Tb)

| Group | No. of Participants | Mean Total ±SD Time (Ta) | Mean Total ±SD Time (Tb) | t-test (p-value) | Productivity Gain (mins.) | Productivity Gain (%) |
|---|---|---|---|---|---|---|
| **Ga** | 29 | 53± 12.87mins | 43± 8.54 mins. | 3.76 (.001) | 10 mins. | 19% |

### 4.3.2 Productivity Gain in Gb

This section reports on the productivity gain in Gb between Ta in which students used HT and, after 4 weeks of MTPE training, in Tb in which students used MTPE. Based on results drawn from

previous research (e.g., Yang et al. 2020; Jia et al. 2019), it is expected that students will show productivity gain in Tb when compared to Ta.

The paired-samples *t*-test was conducted on students in Gb to estimate their productivity in the pre-test (Ta) and after the 4-week teaching intervention (Tb). The mean decrease in total translation time scores was 15 minutes (Figure 4-8 above, Table 4-10 below). There was a statistically significant decrease in translation pre-test total time spent (M= 46 minutes, SD= 11.58) to post-test total time spent (M= 31 minutes, SD= 10.37), t (31) = 8.01, p < .001). Therefore, the results from Gb show that there was a productivity gain of 33% in Tb when compared to Ta.

Table 4-10 Productivity gain in Ga (Ta and Tb)

| Group | No. of Participants | Mean Total ±SD Time (Ta) | Mean Total ±SD Time (Tb) | t-test (p-value) | Productivity Gain (mins.) | Productivity Gain (%) |
|---|---|---|---|---|---|---|
| **Gb** | 31 | 46±11.58mins. | 31±10.37 mins. | 8.01(<.001) | 15 mins. | 33% |

The aim of running the longitudinal comparisons in both Ga and Gb was stated in the methodological approaches in chapter 3, i.e., to calculate the actual productivity differences between Ta and Tb for both groups, in order to confirm whether students in Ga and Gb have made any improvements in the post-test (Tb) compared to their results in the pre-test (Ta). The results in both sections confirmed that both groups have gained productivity in Tb compared to Ta. Although, the productivity gain in Gb was larger than that of Ga, however, disparity in students results in Ta became a confounding factor that affected the later analyses. As if productivity gain in Gb was confirmed to be larger than that of Ga, we would not be able to tell whether the reason was due to the different translation method used (MTPE) or was the productivity gain a natural outcome of the whole group being faster as proved through the longitudinal comparison. Therefore, repeated measures ANOVA was selected to test the inter-group productivity gain as it takes into account all these different factors.

### 4.3.3 Inter-group Productivity Gain (Comparing Ta and Tb of Ga vs. Gb)

In the two previous sections of the productivity results, I attempted to confirm whether there was a productivity gain for both Ga and Gb in (Ta and Tb) through measuring the total translation time in minutes and running longitudinal comparisons between Ta and Tb of each groups' total task duration. In this section, however, I have run a horizontal comparison to find out whether MTPE has played a role as a factor in affecting the productivity of the experimental group (Gb). The

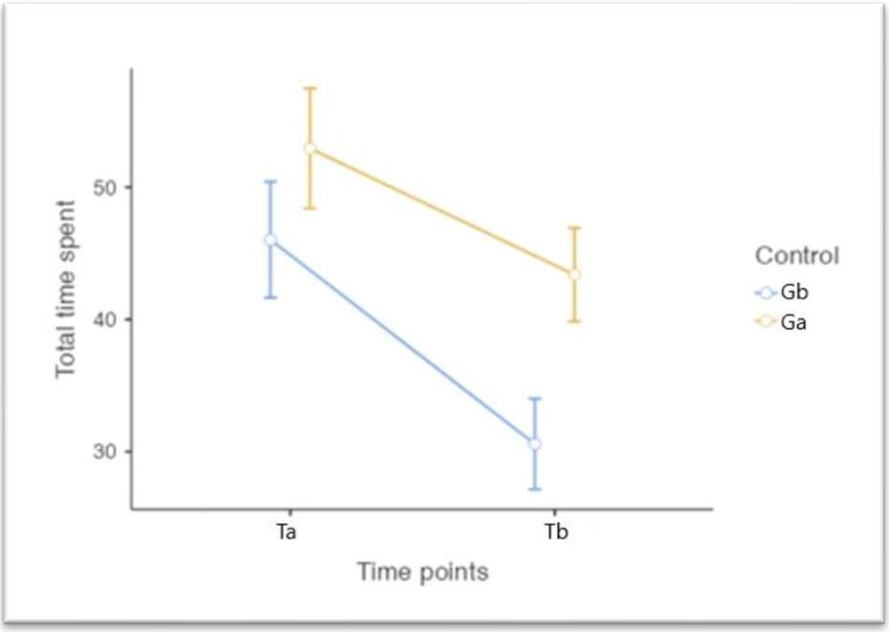
mean duration for Gb in Tb was 31 minutes (M= 31 minutes, SD= 8.54), while the mean duration for Ga in Tb was 43 minutes (M= 43 minutes, SD=10.36). More in-depth analysis found that the productivity gain in Gb was about 28% more than Ga. To compare inter-group productivity gain, repeated measures ANOVA was selected because it: (1) controls for the variation of average total translation time in each group at the two occasions (Phakiti 2015), that is, it takes into account the interaction between time intervals (Ta and Tb) and the average total translation time at each time point. Also, (2) it is robust for deviations when the sample sizes are small (Sullivan et al. 2016). Thus, it represents an advantageous over the traditional *t*-test analysis. Table 4-11 below shows the results of the repeated measures ANOVA.

Table 4-11 Inter-group productivity gain

| | Sum of Squares | df | Mean Square | F | P | $\eta^2_p$ |
|---|---|---|---|---|---|---|
| **TaTb** | 4683.54 | 1 | 4683.54 | 62.48 | < .001 | 0.52 |
| **GaGb ✻ TaTb** | 260.78 | 1 | 260.78 | 3.48 | 0.067 | 0.06 |

Figure 4-9 below visualises the interaction as obtained from the analysis. Although the interaction fell just short of statistical significance at the .05 level, the effect was medium in size [*F* (1, 260.78) = 3.48, p = .06, $\eta^2p$ = .06].

Figure 4-9 Inter-group productivity gain




According to Cohen (1992), an effect size $\eta^2p = .06$ can be considered medium in magnitude (Figure 4-10 below). The medium effect size means that even though the p-value was not significant, however, the effect should not be ignored.

Figure 4-10 Values of effect sizes and their interpretation (Cohen 1992)

| *Kind of Effect Size* | *Small* | *Medium* | *Large* |
|---|---|---|---|
| *r* | .10 | .30 | .50 |
| *d* | 0.20 | 0.50 | 0.80 |
| $\eta^2_p$ | .01 | .06 | .14 |
| $f^2$ | .02 | .15 | .35 |

All in all, the results from running two longitudinal comparisons (Ga between Ta and Tb and Gb between Ta and Tb) showed significant productivity gain in both groups with more productivity gain evident in Gb. Whereas the horizontal comparison of Ga and Gb in Tb showed no significance in terms of the p-value. However, the eta squared ($\eta^2_p$) medium effect size clearly states that MTPE

cannot be ignored as a method of increasing productivity. Further discussion of the productivity results is provided in section 5.3.

## 4.4 Translation Quality Results

This section discusses the translation quality of two groups of translation students (i.e., Ga and Gb), and then the results of each group are compared to one another in an attempt to evaluate the effectiveness of MTPE. The ultimate aim of this section is to attempt to explore the answer to RQ3 through examining the differences in the product between HT and MTPE through error count, and through identifying the error type in both HT and MTPE.

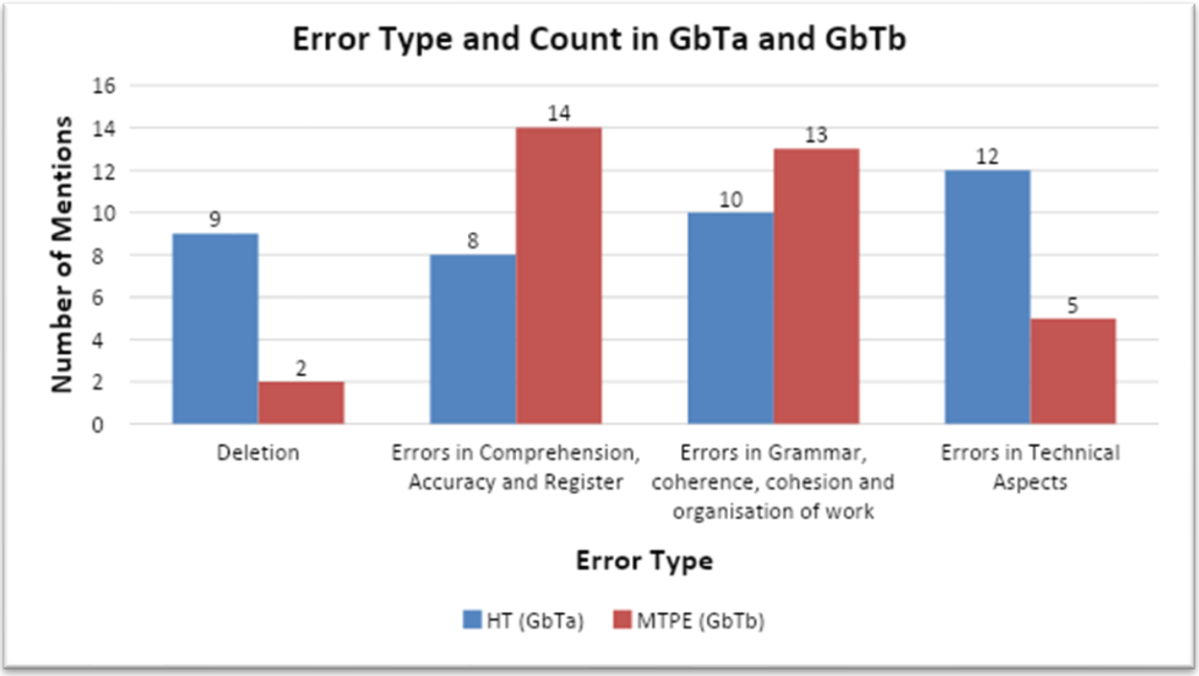
### 4.4.1 Error Type

This section addresses the second sub-question of RQ3 *what are the most common errors in HT and MTPE tasks in the language pair English-Arabic.* Results from this section could provide educational stakeholders with some practical modifications onto MTPE training for it to be considered effective in the classroom especially in the under-researched language pair English-Arabic. In addition, the results are expected to yield some recommendations for the developers of Arabic MT to help them focus on refining the quality of the English-Arabic MT output which might lead to not only Arabic MTPE productivity gain but also improvements in quality as well.

**Errors in HT vs. MTPE Tasks**

Figure 4-11 below lists the various error types that were mentioned in the evaluators' annotations when students in the experimental group Gb first manually translated Ta and then post-edited MT output Tb. Note that the reviewers did not know the method of translation used while reviewing. On Figure 4-11 below, Error Type refers to the four categories of error according to the DipTrans *Examiners' Mark Sheet*, while Number of Mentions refers to the total count that each error was spotted by the researcher in the evaluators' annotation not the total score of that error on the *Examiners' Mark Sheet*:

Figure 4-11 Types and number of errors found in GbTa and GbTb




In the pre-test Ta where students translated using HT, all four types of error were spotted in the evaluators' annotation. Errors in technical aspects were the most problematic (12 mentions) in the HT tasks with evaluators' commenting mostly on spelling mistakes and punctuation errors, followed by errors in Grammar (10), then deletion (9), and finally errors in comprehension, accuracy and register (8).

In Tb, MTPE seems to have improved two aspects of translation, i.e., deletion and technical aspects. Deletion was spotted twice in the evaluators' annotation for Tb with no student failing the test due to missing 5% or more of the original text. In addition, MTPE seems to have maintained or improved the technical aspects on the TTs as the times errors in technical aspects were spotted dropped from 12 times in Ta to 5 times in Tb.

However, the number of errors in AoPs 1 and 3, i.e., in comprehension, accuracy, and register as well as errors in grammar, coherence, cohesion and organization of work have increased when MTPE was the method of translation used.

Some examples of the different error types in HT can be noted in the annotations below:

> **Deletion:** (Wrench) and (try wrapping an old rag around the connecting nut). The translation has missed a couple of significant elements in the source text which will definitely impact the understanding of the instructions. **AoP2:** The text should be well organised to make it easier for the reader. **AoP3:** Leaving a space before the full stop is a technical error. This is repeated and should be

avoided. Another technical issue is the line break or what is known as the paragraph i.e., start a new idea in a new paragraph.

This translation shows an inadequate grasp of the informational content. **AoP1:** There are a number of clumsy or inappropriate renderings, both major and minor inaccuracies, which distort or impair the message at several points. There are also serious deletions (more than 5%), some incorrect choice of register and terminology, and unidiomatic use of language.

This is an excellent translation with no technical issues. However, it would have been better if some translationese words have been avoided to make this translation reads like an original text. For example: She used the word (خاصتك) which is a literal translation of the word (yours) that is not used in the Arabic language (word-attached pronouns are used for this function).

Fail. Many reasons but mainly due to incoherent text and for missing the **core idea**. A showerhead is not a bathroom tap.

Some examples of the different error types in MTPE can be seen in the annotations below:

Except one or two awkward renderings/structures, the translation shows an excellent command of the subject matter with good transfer of information and evidence of complete comprehension throughout.

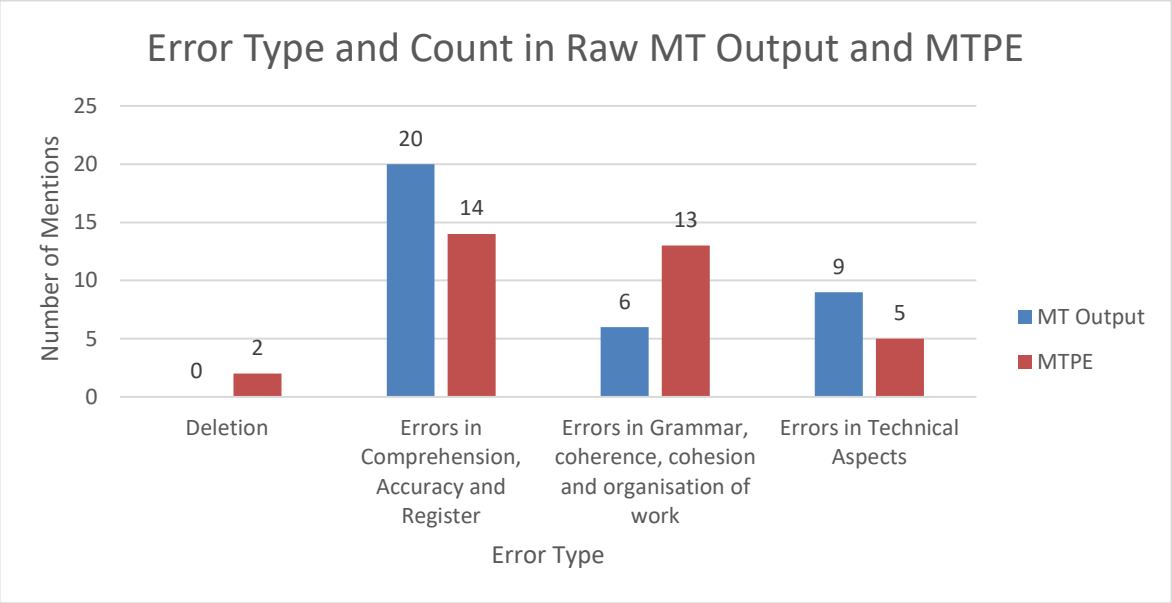
Incoherent text. Includes many major comprehension errors and repetitions. Sounds like an automated translation.

Consistency is a key element in translation. Either use أصلي *''Asli- the Arabic translation for the word Original''* for both words or أوريجينيال *''Original-a transliteration of the English word into Arabic''* for both but not one each.

Except one or two awkward structures, the translation reads like original with an excellent grasp of content.

In order to decide whether the errors spotted in MTPE were a result of MT output or PE practice, I ran an evaluation of raw MT noting which errors students found, did not find, and their overcorrections.

Figure 4-12 Error Type and Count in Raw MT Output and MTPE




The results in figure 4-12 show the types and count of errors found in raw MT output as well as those found in MTPE, suggesting that errors in comprehension, accuracy and register contributed the most errors in both outputs. The results indicate that students managed to decrease errors in comprehension, accuracy and register as well as errors in technical aspects. However, the number of errors in grammar, coherence, cohesion and organisation of work increased in MTPE which indicates a tendency to overcorrect using wrong forms of grammar or lack of cohesion and coherence in the final TT. Deletion was also a result of students' interference with the MT output as it was spotted twice in MTPE but not in raw MT output.

All in all, regarding error types in MTPE tasks, errors in comprehension, accuracy and register as well as errors in grammar, coherence, cohesion and organisation of work mentioned in Tb were more frequent than those in Ta, which indicate that the MT quality of Arabic language is lacking improvement, or that the PE skills the students were taught during the 4-week course were not practiced enough to have a significant impact on the quality of the TTs. Therefore, due to the fact the chosen texts for Ta and Tb were of comparable difficulty, it seems legitimate to conclude that Gb score differences between Ta and Tb were not caused by the text type, but rather they can be attributed to the different method of translation used.

In section 5.4.1, a discussion of the results is provided. The list of limitations and recommendation for future research are provided later in chapter 6.

### 4.4.2 Error Count

In this section, the mean score of each group was compared both horizontally and longitudinally to examine the difference in the overall quality between the product of HT and the product of MTPE. The next two sections (i.e., horizontal analysis and longitudinal analysis) show the comparisons of the results of tests Ta and Tb between groups Ga and Gb.

#### 4.4.2.1 Horizontal Analysis

The horizontal analysis is undertaken to contrast, on the one hand, the scores between the results of the two groups in the pre-test (i.e., GaTa and GbTa) and, on the other hand, the scores between the results of the two groups in the post-test (i.e., GaTb and GbTb). The aim of the horizontal analysis of Ta is twofold: (1) to calculate the baseline, or to decide on the translation ability levels of the two groups before the training, and (2) to collect the baseline scores against which the scores of the post-test would be compared later. Whereas the aim of the horizontal analysis of Tb is carried out to evaluate the effect of MTPE on the performance by comparing the scores of students in Gb against the scores of students in Ga (See Table 4-12 below). The comparisons are carried out by considering the mean of final scores (fs) obtained by all students in each group.

Table 4-12 The aims of horizontal analysis

<table>
<tr><th colspan="2">Ta</th><th>Tb</th></tr>
<tr><td rowspan="2">Horizontal Analysis<br>1. To decide the translation abilities of the 2 groups,<br>2. To collect the baseline scores.</td><td>Ga</td><td rowspan="2">Horizontal Analysis<br>To evaluate the effect of MTPE on Gb compared to Ga.</td></tr>
<tr><td>Gb</td></tr>
</table>

#### 4.4.2.1.1 Horizontal Ta: Baseline Calculation

Based on the technique of cluster sampling that was used for recruitment in the current study, it was assumed that students in Ga and Gb will have similar performance in Ta because I supposed that they shared a similar education level, similar background, and have both been receiving the same form of conventional teaching of translation. In order to consolidate and testify this assumption, a comparison was carried out between the pre-test final scores Ta (fs) of students in

Ga and Gb. Table 4-13 below lists the students' IDs for both groups, their corresponding figure numbers, and the final score (fs) obtained by each student.

As it can be seen in Table 4-13 below, the mean value of GbTa is 74 points which is 4 points lower than GaTa (mean=78 points); a result which deviates from the initial supposition that Ga and Gb will have a similar performance in Ta.

Table 4-13 Ta (fs) of Ga and Gb

| Figure No. | Group A ID | GaTa-fs | GbTa-fs | Group B ID | Figure No. |
|---|---|---|---|---|---|
| 1 | GAN11 | 59 | 49 | GBN18 | 1 |
| 2 | GAN21 | 65 | 55 | GBN04 | 2 |
| 3 | GAN01 | 67 | 57 | GBN29 | 3 |
| 4 | GAN10 | 72 | 61 | GBN13 | 4 |
| 5 | GAN22 | 72 | 64 | GBN27 | 5 |
| 6 | GAN27 | 72 | 64 | GBN31 | 6 |
| 7 | GAN29 | 72 | 68 | GBN19 | 7 |
| 8 | GAN26 | 74 | 71 | GBN22 | 8 |
| 9 | GAN03 | 75 | 71 | GBN30 | 9 |
| 10 | GAN16 | 75 | 72 | GBN03 | 10 |
| 11 | GAN20 | 75 | 72 | GBN11 | 11 |
| 12 | GAN02 | 76 | 72 | GBN15 | 12 |
| 13 | GAN05 | 76 | 72 | GBN20 | 13 |
| 14 | GAN08 | 76 | 74 | GBN16 | 14 |
| 15 | GAN14 | 76 | 74 | GBN28 | 15 |
| 16 | GAN23 | 77 | 76 | GBN12 | 16 |
| 17 | GAN15 | 79 | 76 | GBN17 | 17 |
| 18 | GAN17 | 80 | 76 | GBN24 | 18 |
| 19 | GAN12 | 81 | 78 | GBN25 | 19 |
| 20 | GAN04 | 82 | 79 | GBN02 | 20 |
| 21 | GAN09 | 84 | 79 | GBN07 | 21 |
| 22 | GAN18 | 84 | 79 | GBN09 | 22 |
| 23 | GAN06 | 86 | 79 | GBN10 | 23 |
| 24 | GAN24 | 86 | 80 | GBN01 | 24 |
| 25 | GAN28 | 87 | 80 | GBN05 | 25 |
| 26 | GAN07 | 88 | 80 | GBN14 | 26 |
| 27 | GAN19 | 88 | 80 | GBN23 | 27 |
| 28 | GAN25 | 88 | 81 | GBN08 | 28 |
| 29 | GAN13 | 90 | 87 | GBN26 | 29 |
| | | | 90 | GBN06 | 30 |
| | | | 96 | GBN21 | 31 |
| | Mean | 78 | 74 | Mean | |

Noting that the pass score is 60 points, while one student from Ga failed Ta (scored 59), three students from Gb failed it, with the scores of 49, 55, and 57 points, respectively. These scores may be partially responsible for the lower GbTa (fs) mean. In an attempt to explore the reasons behind the weak performance of these three students, I turned to the detailed *Examiners' Mark Sheet* allocated for each student. According to the two academics who evaluated the students' TTs, students in Gb who failed the translation test performed badly in one or more Aspect of Performance (AoP). The evaluators provided the following comments when annotating the student marking sheet:

**GbN04 (fs = 55)**

> 'The translation achieved some points correctly. However, it needs some improvement at all Aspects of Performance (comprehension and accuracy, grammar, and technical aspects). The most significant advice would be to pay attention to readability of the text in Arabic. Some points do not seem to be Arabic at all, for example: وقاء زجاجي لين.'

**GbN18 (fs = 49)**

> 'The translation is poor at both comprehension, accuracy and register (AoP1) and technical levels (AoP3).'

**GbN29 (fs = 57)**

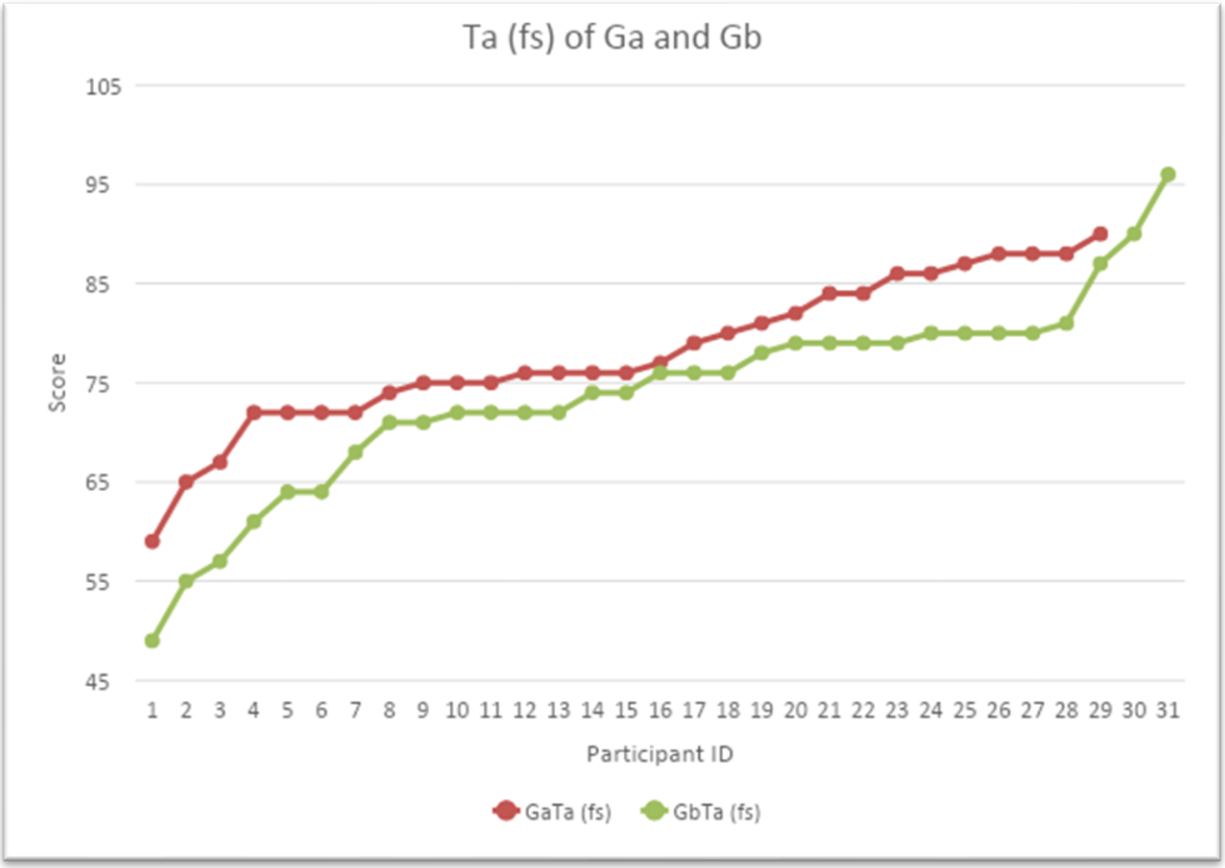
> 'Unfortunately, this translation failed because it misses a whole line from the source text (5% or more) (An old toothbrush) and other words such as "White and Small". Apart from the missing part, it sounds like an adequate translation with some stylistic and grammatical issues.'

In regard to performance of students in Ga, the evaluators commented that **GaN11 (fs = 59)** failed the translation task because she has not completed translating the whole text:

> 'Fail: missing 5% or more of the ST. See annotation.'

On the other hand, when comparing the translation mark sheets of students who performed the best in Ta (i.e., GaN13, GbN06 and GbN21), I found that almost all students had strong command of AoP1 and AoP3 (comprehension, accuracy, register, and technical aspects) with minor errors in AoP2 (i.e., grammar, cohesion and coherence). According to the evaluators, the main difference between the students is that GbN21 had performed better in AoP2 (grammar, cohesion and coherence), while GaN13 and GbN06 had lost a few points due to structuring and linkage errors. The following Figure 4-12 shows the (fs) values of Ga and Gb after having finished Ta.

Figure 4-13 Ta(fs) of both Ga and Gb




Almost all the GbTa (fs) values lie under the GaTa (fs) values, with only two exceptions: Figure No. 30 and 31. As explained in Table 4-13 above, Figure No. 30 is assigned to candidate GbN06, and Figure No. 31 is assigned to candidate GbN21. GbN06 has a score of 90 and GbN21 of 96 in (Ta). The results of the two groups in the Figure numbers ranging between No. 10 and No. 20 seem to be the closest, which suggests that their educational background and personal experiences may also be close. When examining the Examiners' Mark Sheets of candidates GbN06 and GbN21, I noticed that both students had a strong command of AoP1 and AoP3 (comprehension, accuracy, register, and technical aspects) with minor errors in structuring and linking.

Indeed, the Ta (fs) of both Ga and Gb demonstrated in Figure 4-12 above provided a baseline for comparison, however, the scores showed that disparity exists between the two groups with Gb performing worse. Although the results of all students in both groups showed errors in all AoP, the overall lower mean final scores of Gb indicate that students in this group did not know how to deal with comprehension, accuracy, and technical aspects of the text as well as students in Ga did. In the end, it is important to note that although randomization is considered successful when

group results are comparable at the baseline level so as to ensure confidence that the observed difference in the outcomes between the groups is a result of the teaching intervention rather than the confounding factors (De Boer et al. 2015), this variation of baseline results seems unavoidable because of the sampling criteria used to fulfil the purpose of the study which aimed at evaluating the effectiveness of MTPE in an unadjusted natural group of students (See Cluster Sampling: Saldanha and O'Brien 2014: 34).

#### 4.4.2.1.2 Horizontal Tb: Inter-group MTPE Effectiveness

This part of the study assumed that all students in both Ga and Gb would obtain a higher score in Tb than Ta, simply because they had been trained in translation for four more weeks, as part of their BA degree, and regardless of whether Gb had been trained in the use of MTPE or not. The aim of the horizontal comparison between the two groups in Tb is to evaluate the overall effect of MTPE on Gb compared to Ga.

As illustrated in Table 4-14 below, the mean value of GaTb is 83 which is 5 points higher than the mean value of 78 for the same group in Ta, whilst the mean value of GbTb is 80, six points higher than the mean value of 74 points that they obtained in Ta. These results clearly support the assumption that both groups of students have performed better in the Tb than in Ta.

Table 4-14 Tb (fs) of Ga and Gb

| Figure No. | Group A ID | GaTb-fs | GbTb-fs | Group B ID | Figure No. |
|---|---|---|---|---|---|
| 1 | GAN11 | 68 | 60 | GBN10 | 1 |
| 2 | GAN22 | 75 | 73 | GBN04 | 2 |
| 3 | GAN08 | 75 | 73 | GBN22 | 3 |
| 4 | GAN21 | 76 | 74 | GBN03 | 4 |
| 5 | GAN27 | 78 | 74 | GBN29 | 5 |
| 6 | GAN14 | 78 | 74 | GBN28 | 6 |
| 7 | GAN15 | 80 | 74 | GBN24 | 7 |
| 8 | GAN01 | 81 | 75 | GBN18 | 8 |
| 9 | GAN12 | 81 | 75 | GBN16 | 9 |
| 10 | GAN16 | 82 | 77 | GBN31 | 10 |
| 11 | GAN06 | 82 | 77 | GBN07 | 11 |
| 12 | GAN07 | 82 | 78 | GBN15 | 12 |
| 13 | GAN10 | 83 | 78 | GBN17 | 13 |
| 14 | GAN26 | 84 | 78 | GBN25 | 14 |
| 15 | GAN02 | 84 | 79 | GBN30 | 15 |
| 16 | GAN05 | 84 | 79 | GBN01 | 16 |
| 17 | GAN09 | 84 | 81 | GBN12 | 17 |
| 18 | GAN03 | 85 | 81 | GBN23 | 18 |
| 19 | GAN25 | 85 | 82 | GBN27 | 19 |
| 20 | GAN04 | 86 | 82 | GBN26 | 20 |
| 21 | GAN29 | 87 | 83 | GBN05 | 21 |
| 22 | GAN17 | 87 | 83 | GBN08 | 22 |
| 23 | GAN13 | 87 | 84 | GBN19 | 23 |
| 24 | GAN20 | 88 | 84 | GBN06 | 24 |
| 25 | GAN24 | 88 | 84 | GBN21 | 25 |
| 26 | GAN23 | 90 | 85 | GBN13 | 26 |
| 27 | GAN28 | 90 | 85 | GBN11 | 27 |
| 28 | GAN18 | 91 | 86 | GBN20 | 28 |
| 29 | GAN19 | 92 | 89 | GBN02 | 29 |
| | | | 90 | GBN14 | 30 |
| | | | 96 | GBN09 | 31 |
| | Mean | 83 | 80 | Mean | |

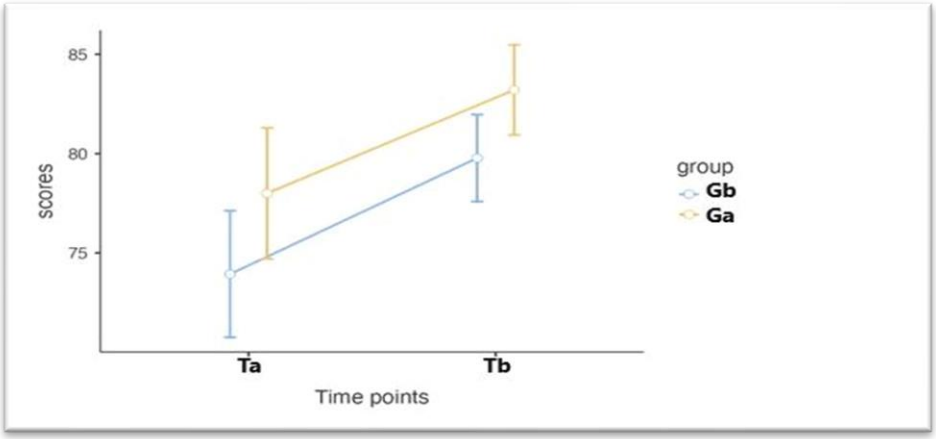
Repeated measures ANOVA was selected to answer this question because it controls for the variation of average scores in each group (Ga and Gb) at the two occasions (Ta and Tb) (Phakiti 2015). That is, it takes into account the interaction between the time (Ta and Tb) and average scores of each group at each time point. Figure 4.13 below illustrates the interaction as obtained from the repeated measures ANOVA. As it can be seen in the figure, the interaction shows very similar trend of improvement across Ga and Gb, the interaction is statistically non-significant and trivial in size [*F* (1, 2.99) = 0.09, p = .77, $\eta^2p$ = .00] (Table 4-15 below).

Table 4-15 Inter-group MTPE effectiveness

| | Sum of Squares | df | Mean Square | F | p | $\eta^2_p$ |
|---|---|---|---|---|---|---|
| **TaTb ✻ GaGb** | 2.99 | 1 | 2.99 | 0.09 | 0.772 | 0.00 |

However, since the scores of both groups have increased in in a rather similar manner (Figure 4-14 below), we can conclude that MTPE is effective in achieving similar scores when compared to HT. This is being said regardless of the fact that GaTb final scores mean (the yellow line) lies above that obtained by Gb which suggests that Ga has achieved better overall scores in terms of translation quality in both tests. Answers to the question about whether *TTs resulting from MTPE show similar pattern of errors as those resulting from HT* can be found in section 4.4.1.

Figure 4-14 (fs) of Ga and Gb in both Ta and Tb




In summary, we can confirm that the results of the translation quality score comparison support the assumption that students who use MTPE would show similar results when compared with students who use HT in terms of translation quality scores in the post-test. However, neither the

size of the p-value nor the effect size confirmed that there is evidence of a difference between the quality of Gb and Ga in favour of the first which can be interpreted as the failure of MTPE in getting students who used MTPE in Gb to perform better than students who used HT in Ga. A further discussion of these results is provided in 5.4.2.

#### 4.4.2.2 Longitudinal Gb: Intra-group MTPE Effectiveness

As explained in section 3.9.4, the purpose of the longitudinal analysis is to evaluate the effectiveness of MTPE on the translation quality scores of the same group of students (Gb). This is done through calculating the final score (fs) differences of Tb minus Ta for the experimental group Gb, so as to ascertain the improvements accomplished by the students as a group unit. To achieve this aim, this section reports on the longitudinal analysis of the effect of MTPE (if any) on the translation quality of Gb after the 4-week training course through the comparison of final scores (fs). The final scores (fs) before the intervention (GbTa (fs)) were deducted from the final scores (fs) after the intervention (GbTb (fs)), and the significance of the differences in scores was calculated to confirm the improvement or deterioration in translation quality related to MTPE. The disparity of pre-test results of the two groups of students (Ga and Gb) have added to the importance of running this longitudinal analysis. This within-Gb comparison is specifically vital because the significance or non-significance of the intervention (PE) would prove more valid when all the variables are unified.

The first step consists of comparing the change in final scores (fs) between GbTb (post-test) and GbTa (pre-test) where scores in Ta are subtracted from those in Tb (i.e., Gb (fs) Tb-Ta). This result will help ascertain whether Gb has made overall improvements in the translation quality. If this calculation resulted in a positive value (+), this indicates that students in Gb have scored better in Tb when they used MTPE whereas if the results from the calculation showed negative values (-), this indicates that students in Gb have scored worse in Tb when they used MTPE.

Then, the second step involves calculating the significance (p-value) of the difference in the final scores of Gb between Ta and Tb (Gb (fs) Tb-Ta) to ensure whether the difference is significant.

Table 4-16 below offers a synopsis of the two questions to be answered in this section together with the corresponding methods of analysis employed to find the various answers:

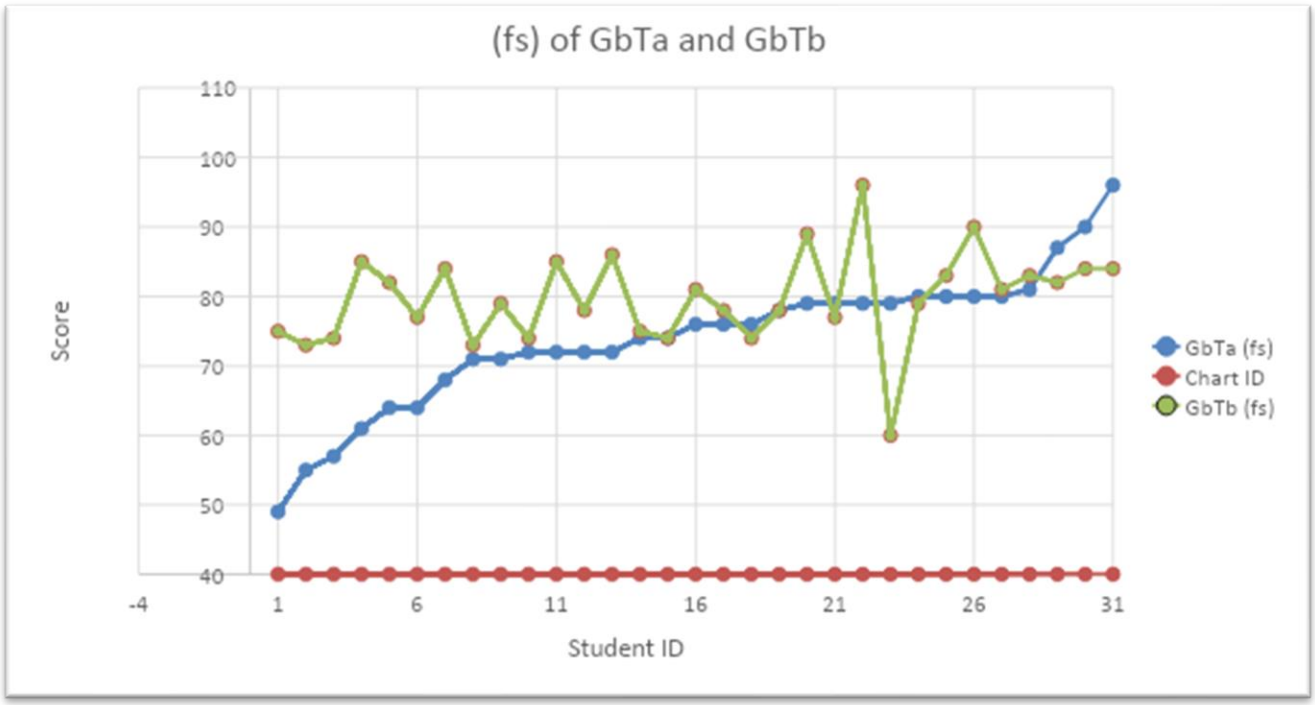
Table 4-16 Research questions and methods

| Research Question | Method of Analysis to Find the Answer |
|---|---|
| 1.Has Gb made any improvements in terms of the fs? | Calculate (Gb (fs) Tb-Ta). |
| 2. Is the improvement made by Gb significantly different in terms of the fs? | Calculate p-value for (Gb (fs) Tb-Ta). |

The longitudinal analysis in the current section looks at the comparison between the final scores obtained by Gb students in Ta and Tb tests. Theoretically, any potential disparity between the final scores (fs) achieved in Ta and Tb will indicate the degree of overall quality improvement or deterioration experienced by the group after the four weeks' training.

Figure 4-15 below presents the ID numbers of students in Gb, their final scores (fs), and the score difference between the two tests ((fs) Tb-Ta).

Figure 4-15 (fs) of GbTa and GbTb




As illustrated in Figure 4-15 above, 22 students out of 31 students who comprise Gb have improved their performance, 2 students performed the same, and 7 students performed worse after 4 weeks' training.

An interesting pattern appears on Figure 4-14 above where students who performed the worst in Ta (Chart IDs 1 through 13 except 8 and 10) made the overall best advantage of using MTPE in Tb, whereas the students who performed worse in Tb (chart IDs 29, 30, and 31) were the ones who achieved the best scores in Ta.

The students who made the most significant improvement include chart ID numbers 1, 4, 5, and 22 (GbN18 and GbN13, GbN27, and GbN09), with increased scores of 26, 24, 18, and 17 points respectively, whereas the least improvement, apart from students who scored the same as their Ta scores or less, corresponds to GbN16 and GbN23, who only increased their score by 1 point.

Given the unexpected performance by students GbN10 and GbN21, with respectively negative Tb-Ta (fs) of -19 and -12, I revisited their task sheets, their examiners' mark sheets and their time counts to try to find an explanation. GbN10 (-19 points) spent 37 minutes in finishing Ta and spent 30 minutes in Tb which indicates a reasonable productivity gain, but this gain in productivity did not reflect an improvement in quality. The evaluators' comments on this student's performance in Ta was: '*The translation shows a good command of the subject matter, although at times there are some grammar and linkage mistakes.*', in addition, the student's scores in AoP1 and AoP3 were relatively high indicating that the student did well in terms of comprehension, accuracy and technical aspects in Ta. However, in Tb, the reviewers' comment was: '*The translation includes a number of clumsy or inappropriate renderings, some awkwardness, and lapses in grammar*.' This comment, in addition to the total scores of the students, indicate that the performance has been seriously affected in terms of comprehension, accuracy as well as grammar when the student was post-editing MT output. When I checked the Tb task of the student, I noticed that she has copied and pasted the MT output correcting minimal errors although the instructions were clear that the highest possible quality was required when they post-edited the MT output.

Student GbN21 (-12), on the other hand, spent 55 minutes in finishing Ta and 39 minutes in Tb which again indicates a productivity gain in favour of quality. The evaluators' comment on this student's performance in Ta was: '*Adequate command of the subject matter with minor errors in structuring and linking*.' However, in Tb, the evaluators' comment was: '*Translation is acceptable, though, with some errors in understanding, reformulating, linking and structuring*.' This indicates, again, that the performance has been affected in terms of comprehension, accuracy as well as grammar when the student was post-editing MT output. When I checked the Tb task of the student, I noticed that she has performed minimal PE on the MT output leaving behind plenty of literal renderings of vocabulary, wrong Arabic pronouns, and spelling mistakes. In both cases, full post-editing of MT output was not performed by the students, and the only apparent positive effect of MTPE was the gain in productivity.

Table 4-17 below shows (fs) obtained by Gb students in Ta and Tb. As can be concluded, most of Gb (22 students) scored higher in Tb than in Ta except for 7 students, whose translation quality deteriorated while two students' performance remained the same. Generally speaking, the students in the lowest score zone in Ta made on average more improvement than the students in the middle and the high range.

Table 4-17 Final scores of GbTa and GbTb

| Group B ID | Chart ID | Gb-Ta-fs | Gb-Tb-fs | (Tb-Ta) fs |
|---|---|---|---|---|
| **GBN18** | 1 | 49 | 75 | **+26** |
| **GBN04** | 2 | 55 | 73 | **+18** |
| **GBN29** | 3 | 57 | 74 | **+17** |
| **GBN13** | 4 | 61 | 85 | **+24** |
| **GBN27** | 5 | 64 | 82 | **+18** |
| **GBN31** | 6 | 64 | 77 | **+13** |
| **GBN19** | 7 | 68 | 84 | **+16** |
| **GBN22** | 8 | 71 | 73 | **+2** |
| **GBN30** | 9 | 71 | 79 | **+8** |
| **GBN03** | 10 | 72 | 74 | **+2** |
| **GBN11** | 11 | 72 | 85 | **+13** |
| **GBN15** | 12 | 72 | 78 | **+6** |
| **GBN20** | 13 | 72 | 86 | **+14** |
| **GBN16** | 14 | 74 | 75 | **+1** |
| **GBN28** | 15 | 74 | 74 | **0** |
| **GBN12** | 16 | 76 | 81 | **+5** |
| **GBN17** | 17 | 76 | 78 | **+2** |
| **GBN24** | 18 | 76 | 74 | **-2** |
| **GBN25** | 19 | 78 | 78 | **0** |
| **GBN02** | 20 | 79 | 89 | **+10** |
| **GBN07** | 21 | 79 | 77 | **-2** |
| **GBN09** | 22 | 79 | 96 | **+17** |
| **GBN10** | 23 | 79 | 60 | **-19** |
| **GBN01** | 24 | 80 | 79 | **-1** |
| **GBN05** | 25 | 80 | 83 | **+3** |
| **GBN14** | 26 | 80 | 90 | **+10** |
| **GBN23** | 27 | 80 | 81 | **+1** |
| **GBN08** | 28 | 81 | 83 | **+2** |
| **GBN26** | 29 | 87 | 82 | **-5** |
| **GBN06** | 30 | 90 | 84 | **-6** |
| **GBN21** | 31 | 96 | 84 | **-12** |

Therefore, to answer the second question about whether the improvement made by Gb was significantly different in terms of the (fs), a paired-sample *t*-test was conducted to evaluate the impact of MTPE on Gb students' translation scores (Gb (fs) Tb-Ta). The mean increase in total scores was 5.84 points with a 95% confidence interval ranging from 2.09 to 9.59. There was a statistically significant increase in the scores from Ta (M= 73.9 points, SD= 9.9) to Tb (M= 79.8 points, SD= 6.6), t (31) = 3.78, p = .003) (See Table 4-18 below).

Table 4-18 Translation quality scores (Ta vs. Tb)

| Group | No. of Participants | Mean fs ±SD (Ta) | Mean fs ±SD (Tb) | t-test (p-value) | Score gain (95%CI) |
|---|---|---|---|---|---|
| **Gb** | 31 | 73.9±9.9min. | 79.8±6.6 mins. | 3.78(.003) | 5.89 (95%CI: 2.09 to 9.59) |

A discussion of all the results from the longitudinal analysis is provided in section 5.4.2.

# Chapter 5 Discussion

## 5.1 Introduction

In the previous chapters, I reported results from exploring students' opinions through FGDs. In addition, retrospective pre-test survey responses were examined to triangulate the findings from the FGDs and to determine whether there was a change in students' opinions after the MTPE intervention (RQ1). I also reported statistical results in relation to the research question about productivity (RQ2) after comparing students' translation task times between those who used HT in both the pre-test and the post-test and those who used HT then MTPE. Finally, I reported statistical and descriptive results to answer the final research question (RQ3) about translation quality through analyzing error count and error type. In this chapter, I offer summaries of the results and their discussion.

This chapter is divided into five sections including the current introduction. The second section (5.2) discusses the qualitative findings from the FGDs and the quantitative results from the retrospective pre-test survey. The third section (5.3) discusses the results from the productivity experiment. The fourth section (5.4) is composed of two parts: the first (5.4.1) identifies the error types in both methods of translation, and the second (5.4.2) discusses the results from the comparison of error count between HT and MTPE. Finally, the fifth section (5.5) provides a summary of the discussions provided in this chapter.

## 5.2 Students' Opinions about MTPE

This section discusses the degree of reward, usefulness, ease of use, productivity, quality, and the preferred method of translation by students who took part in the FGDs and, later, by those who comprised the experimental group (Gb) who took the retrospective pre-test survey.

**RQ1: What are the differences in students' opinions about HT and PE?**

Before commencing a task, the person has to show a positive attitude towards it in order to succeed and this observation applies to using MTPE (Daems 2016). However, previous research showed that participants did not always show positive attitudes to MTPE all the time (e.g., Al-Mutawa and Izwaini 2015; Moorkens and O'Brien 2013; Moorkens and O'Brien 2015). In order to evaluate the effectiveness of MTPE based on examining translation students' opinions in the current study, the aim of RQ1 was to explore students' opinions about MTPE training and use through comparing their opinions about it with their opinions about HT which is an established

practice in the classroom. For the researcher to examine whether a translator reveals positive or negative attitude towards MTPE, a number of criteria were set by Daems (2016) to help explore whether translators accept or reject MTPE. This section will list these criteria for the language pair English-Arabic for Native Arab female translation students. It is worth mentioning that the scope of the criteria representing translation students' opinions about HT and MTPE is limited to the results of the present study.

The overall findings from examining the FGDs and the survey responses of the translation students are summarized in the following section. Later, all discussions of the sub-questions that comprise the overall answer to RQ1 are provided.

### 5.2.1 Discussion of Findings from FGDs and Retrospective Pre-test Survey

From the comparison of findings from the FGDs and survey, it can be tentatively concluded that students' opinions that were revealed through the FGDs were generally mixed with a bias in favor of HT, except for speed regarding which most of the students thought that MTPE was the faster method of translation. As for the survey, the results from 'before the intervention' (hereafter, the before-survey) support those revealed in the FGDs. However, the 'after the intervention' (hereafter, the after-survey) responses revealed a statistically significant shift towards more acceptance of MTPE training and use indicating that the more students learned about the features of MT and PE skills, the better their opinions became (in line with Alotaibi 2014; Çetiner 2018). The overall results also corroborate Daems' suggestion that 'understanding indeed leads to acceptance' (2016: 162). Therefore, the results from the current study suggest that introducing MTPE skills into the curriculum of undergraduate English-Arabic translation programmes may be a step forward. The following six sections provide further discussions about the sub-questions of RQ1.

#### 5.2.1.1 Degree of Reward

**How rewarding is PE compared to HT?**

The findings from the FGDs revealed that students were split between HT and MTPE with more biased opinions towards HT and against MTPE. Those who thought MTPE is more rewarding listed the following reasons that underpin their opinions: MTPE is faster than HT, clients generally do not care about the quality, but they want a fast service, MT output can be edited, and MT output is rapidly improving. On the other hand, those who thought HT is more rewarding thought that it is so because: HT quality is better, MT jeopardizes job opportunities, MT output is full of errors, HT brings about better translation style, and one student thought that HT is faster than MTPE.

The overall findings from the FGDs generally support those of Guerberof-Arenas (2013), who found that translators have mixed experiences and feelings about MT output and PE. However, although the before-survey responses corroborate those of Daems (2016) who found that translators thought HT to be more rewarding than MTPE, the after-survey responses revealed a positive shift in students' opinions in favour of MTPE. A possible explanation is that because MTPE is not incorporated in the teaching of translation in this programme, students' opinions about MT and PE were stemming from their personal experiences and preferences. It could also be related to a lack of trust in technology or a lack of knowledge in terms of what translation software is capable of. In addition, some students who revealed that they did not think MT was rewarding showed not only linguistic rejection but also psychological rejection. In line with Krings (2001) and Loffler-Laurian (1996), some students declared that they feared (or anticipated) that the machine would replace them in the job market, and therefore, that is why they refused embracing it as a rewarding method of translation. As a suggestion, translation students need to browse job advertisements which seek MTPE expertise so that they become aware that MT and PE are skills required of professional translators. Another explanation for the seemingly biased attitudes against MTPE whether in the FGDs or in the before-survey responses may be attributed to the notion of the 'learned attitude' (Doherty and Moorkens 2013: 132) as teachers still consider using MT as an act of cheating when completing translation tasks and assignments in the college where this study took place (Al-Jarf 2017). Rather than advising against MT use, teachers are encouraged to highlight the ethics of translation as a profession.

On the other hand, results from the after-survey showed that the intervention was successful in demonstrating the degree of reward of PE in the daily life of a translator as the after-survey results showed that there is a significant positive change in students' opinions. Students' opinions after the intervention support those of Daems (2016), who found that students' feelings seem somewhat more positive towards PE after the translation task which involved using PE skills. Also, the after-survey results support those of Alotaibi (2014) and Çetiner (2018), who ran pre-post studies. In their post-surveys, students showed more acceptance towards the incorporation of technology in the translation process and CAT tools in general.

Doherty and Moorkens (2013) concluded that if students perceive a skill to be of value, they will be motivated to learn it, while if negative attitudes exist towards it, then students will be unlikely to want to learn the skill. In the current study, although students stated that they often/always used MT, their motivation to embrace it was poor because of the disappointing errors in the MT output, and because of their fear that it might risk their job opportunities. Students' poor motivation was validated through the students' tendency to consider MTPE not rewarding. Students' association between MTPE to being rewarding and the errors in the output supports

Guerberof Arenas's discussion in which a professional translator thought that if MT output was 'understandable' (2013: 86), that they would consider it financially rewarding. Therefore, I would recommend that classroom training highlights the strengths and weaknesses of MTPE on the overall productivity and quality of translation which in turn may increase students' chances in the translation job market.

#### 5.2.1.2 Usefulness

**How useful is MT output according to translation students?**

Again, students' responses in the FGDs revealed that they were split between thinking that MT output is 'useful', and that MT output is 'less useful'. This result does not support that of Daems (2016) in which students were more inclined to choose 'MT output is often useful'. However, the results from the survey indicate that when students were given the chance to learn more about the potentials of MT output and how to tackle its weaknesses, their responses became closer to those in Daems' study. This shift in students' opinions towards embracing MT suggests that the more students are exposed to MT while learning how to tackle its shortcomings, the less biased opinions they expressed about it which in turn provides them with better chances to exploit MT to its fullest potential.

In the FGDs, students highlighted that MT output can be considered useful because it promotes faster translation jobs especially in the fields where fixed terms are dominant such as in technical and legal translation. This is a result that I considered interesting since it mainly stemmed from students' personal experiences with MT as they are not provided with MT training that suggests which text types are suitable for MT. This finding supports Reinke's (2018) claim that MT is only suited for a limited range of text types for use through an MTPE process. However, a student in Gb (GbN04) who participated in the teaching intervention revealed an extreme opinion in favour of trusting the usefulness of MT output with all text types to the extent that she simply assumed all MT-generated terms are accurate in the context they appear in. This particular case required deeper investigation because it contradicted the results from Cadwell et al. (2016) where translators lacked the trust in MT output and therefore its usefulness. The most obvious difference between the translators in Cadwell et al.'s study and this student in the level of experience as they were experienced translators in Cadwell et al. Verifying this student's lack of experience was made through checking her task scores (she scored 55 in Ta and 73 in Tb) and her responses about MT use (when asked about MT use, she responded with 'somewhat disagree' in the pre-intervention, and with 'strongly agree' in the post-intervention survey). In other words, based on results from the current study as well as previous studies, it can be suggested that

highlighting the weaknesses of MT output is recommended to maintain the required balance between the translators' degree of trust and their use of MT.

On the other hand, some students regarded MT output as 'not useful' because it is full of errors, fails in respecting the Skopos for the target text, and is unable to translate metaphors nor consider cultural issues. It is possible that the quality of the Arabic MT output has affected students who thought that MT output was not useful because of errors. Zaghouani et al. (2016) categorised errors in Arabic MT as spelling errors (mostly the letters Yaa' "ي" and hamza "ء" in the MT texts), frequent errors in word choice, incorrect derivation or inflection, wrong gender and number agreement, tense assignment, the error of translating proper names into the Arabic language, and punctuation errors where punctuation signs appear in the wrong place. My assumption is supported by the results from Alotaibi (2014), in which students expressed disappointment in MT output being not very useful because of the low-quality output and limited support for the Arabic language. In this study, the intervention focused on fixing the most common errors in Arabic MT (lexical, grammatical, and stylistic) which seems to have worked as a factor in enhancing the students' opinions about MTPE because they were not as intimidated by the errors but rather, they might have felt more confident when dealing with them which was significantly evident in the more positive opinions about the usefulness of MTPE in the after-survey results. However, the inability of MT to translate metaphors and consider cultural differences cannot be said to have been resolved through the results of this study since it has used technical texts which featured explicit language.

On the other hand, according to Davis, the perceived usefulness which leads to acceptance is 'the degree to which a person believes that using technology would enhance his or her job performance' (1989: 320). Results from the after-survey support the notion of promoting MTPE acceptance which entails students finding it useful. The percentage of those who thought MT output as useful quadrupled while the percentage of students perceiving MT output as not useful sharply dropped (see Table 4-8- Summary of Students' Results). Those positively shifted opinions were may have had a positive impact on students' performance in the MTPE task which showed a similar score increase to that of students who used HT although different types of errors were highlighted in the MTPE TTs (see section 5.4.2). This is consistent with the findings from Daems' thesis (2016) in which students found that MT output was 'sometimes' and 'often' useful in the post-survey. These results also corroborate Moorkens et al. findings (2018), where less experienced translators found MT suggestions useful. Thus, it is suggested that translation students are encouraged to examine MTPE translated texts, so they learn the limitations as well as the common linguistic/stylistic errors in order to gain more confidence that leads to acknowledging the usefulness of MTPE. Also, demonstrating PE tasks that show how productivity

increases while maintaining the quality would provide examples that explain how MTPE enhances students' performance.

#### 5.2.1.3 Ease of Use

This section presents findings related to students' opinions about whether MT output is easy to use. The reported results are drawn from the retrospective pre-test survey only as this question was not asked in the FGDs. When applying Davis's (1989) definition of the perceived ease of use which was explained to the participating students, those who find MTPE free of effort can be considered students who find MTPE easy to use. In the after-survey, two-thirds of the students thought that MT output was easy to use. This result corroborates with Daems (2016) in which students considered PE the least tiring method of translation. However, gauging cognitive effort in PE processes and comparing it to that in HT in the language pair English-Arabic (which was not part of the scope of the present study) is a more solid means to confirm or refute this conclusion.

According to the study conducted by Alotaibi (2014), students showed poor opinions and confused knowledge about CAT tools before the intervention, while they engaged more in using those tools and showed more positive attitudes towards them after the intervention. Therefore, translation students in English-Arabic programmes may need to practice post-editing MT output more in the classroom to promote its use and to make it easier for trainees to be able to handle it.

The results from the after-survey showed that this was the case, indicating that the more students spent time learning MTPE and practiced it, the more positive opinions they revealed about its ease of use. Therefore, it is recommended that students spend more time learning and practising MTPE in class to promote its user-friendliness and ease of use.

#### 5.2.1.4 Speed

**Which translation method is perceived as being faster?**

As for the question about students' opinions about speed, in the FGDs, 75% of the students thought that 'MTPE is generally faster than HT' which is an expected result (de Almeida 2013; Haji Sismat 2016). This finding also supports the results of Alhaisoni and Alhaysony (2017) in which students favored the MT system GT because of its ability to translate texts quickly. A possible explanation for students' tendency to think that MTPE is faster when asked in the FGDs before they had the chance to use it is that they based their opinion on the comparison of translating with a pen and paper which is how they performed their translation tasks in the classroom in addition to their general perception of the fast processing of PCs.

However, results from the before-survey revealed a disparity in students' opinions when compared to the findings from the FGDs as the responses from the survey revealed somewhat moderate opinions about the speed of MTPE (37% thought that MTPE is faster than HT while 29.6% thought otherwise) (see Table 4-8- Summary of students' results). The results of the after-survey revealed that not only the percentage of students thinking that 'MTPE is faster' increased (37% to 48.1%), but also those who thought that 'HT is faster' increased as well (29.6% to 44.4%) (see Table 4-8- Summary of students' results). The latter increase in the percentage of students perceiving HT as the faster method of translation supports the expectation of Gaspari et al. (2014) that participants would perceive PE as being more time-consuming than HT. Nevertheless, this result of the survey contradicts the results of Daems (2016) in which most of the participants agreed after completing the PE task that PE is faster than HT. It is possible that the fact that the teaching intervention took only 4 weeks with complex contents, and not plenty of hands-on MTPE practising in the classroom may have affected the students' opinions about MTPE speed. I say this because some students clearly declared in the FGDs that they considered MTPE as the faster method of translation because HT requires more cognitive effort, especially when the translator tends to unnecessarily overthink the meaning of the ST. Whereas, when students in Gb were offered the chance to learn about the skills required for PE and to put those skills into practice, some of them revealed in the survey that HT is faster. It is possible that they thought so because HT did not require as much technical and cognitive effort as MTPE.

On the other hand, 8 of the 9 students who thought that MTPE is faster nevertheless considered HT as the more rewarding method of translation. According to the definition of reward that was provided in section 4.2.1.2, this indicates that fast turnaround may be a neglected value for those students who seek to produce high quality translations. In other words, we can assume that students who care more about the quality of the TTs may not worry too much about MTPE productivity.

Attitudes of translation students towards MTPE speed can be enhanced through learning more about the potentials of PE. In addition, students need to gain confidence to resist the urge to edit everything that seems faulty because a major advantage of MTPE is increasing productivity and saving time (De Almeida and O'Brien 2010; Plitt and Masselot 2010). They can also make use of PE vs. HT productivity practice in the form of weekly assignments where they are given opportunities to compare the time allocated for translating from scratch with post-editing MT output. More recommendations for Arabic MTPE training can be found in section 6.4.1.

#### 5.2.1.5 Translation Quality

**How is the quality of both methods of translation perceived?**

Most of the students in the FGDs did not seem convinced of MTPE quality in comparison to that of HT. They also highlighted that they preferred HT because it can maintain their personal translation styles which corroborates the results of Koehn (2009) where translators revealed taking pride in their work, believing that technology is not capable of producing useful output. As for the quality of current Arabic NMT output, students seem to have to learn a great deal of translation skills (just as suggested by Yamada 2019) in order to fix the numerous errors in the Arabic NMT output that were listed by Ameur et al. (2020).

On the other hand, in the after-survey, the number of students thinking that they managed to produce better quality through post-editing MT output raised from 11.1% to 70.4%. Both those who thought PE quality was worse or similar have decreased respectively from 51.9% to 7.4% ''worse quality'' and from 37% to 22.2% ''similar quality''. This result confirms that of Çetiner (2018) in which students developed a positive attitude regarding the perception of the quality of translations done with translation tools. It seems that although the quality of the Arabic MT output was full of errors (Zaghouani et al. 2016), students who thought that MTPE produces better translation quality may have become aware of lexical, grammatical as well as stylistic errors through the tips and tricks they learned in the teaching intervention. We can only consider students' 'awareness' of the existence of lexical and grammatical errors without confirming students' ability to fix those errors because as the results of the error type showed (section 4.4.1) that the majority of errors spotted by the evaluators were of lexical and grammatical nature.

#### 5.2.1.6 Preference

**Which translation method is the most preferred?**

This section presents findings related to translation students' preferred method of translation. In the FGDs, the vast majority of students preferred HT. Findings from the retrospective pre-test survey validate these results as the before-survey responses revealed that over 85% of students preferred HT. The findings support those of Doherty and Moorkens (2013) in which most of the students were against MT. It seems that the authors' explanation is applicable on the current findings as well. The authors attributed the sceptical and biased attitude towards MT to be either a 'learnt attitude' or a result of students' lack of technological expertise since they were students of humanities. These explanations can be applied to the findings of the current study because in the college where this study took place, students are banned from using MT and are asked to hand-write their translation on paper, and some teachers refer to students who use MT as cheaters (Al-Jarf 2017). In addition, the largest percentage of applicants to the college where the current study took place were students without much prior experience in technology (Al-Jarf 2017).

In addition, these findings support those by Daems (2016), who found that both professional as well as student translators preferred HT, although they did not mind PE. Indeed, the result about students' tendency to prefer HT in the current study which some explained by saying that HT is better because it allows creativity and takes cultural differences into account supports the findings by Moorkens et al. (2018) where participants preferred translating from scratch, mostly due to the freedom to be creative without the constraints of MTPE that can be seen in segment-level segmentation.

However, students' preferences drastically shifted towards preferring MTPE after the intervention. The after-survey responses showed that a little over 81% of those who took the intervention preferred MTPE. This positive shift resonates with those of Alotaibi (2014) and Çetiner (2018) whose findings showed that students who took a semester-long CAT intervention revealed positive acceptance to embracing CAT tools.

The next section discusses the results of translation productivity.

## 5.3 Discussion of Translation Productivity

This section discusses the productivity of MTPE compared to HT. The aim is to evaluate the effectiveness of MTPE through examining its productivity compared with HT which is the current method of teaching in the university where this study took place. It seeks to answer RQ2 about the differences in the process between HT and MTPE by exploring whether the productivity gain from MTPE output is larger than that of HT.

The results (section 4.3 above) broadly support prior studies which compared MTPE with HT (e.g., Plitt and Masselot 2010; Haji Sismat 2016; Yang et al. 2020): Ga students who used HT in Ta and Tb gained productivity in the post-test. Gb students who used HT in Ta and MTPE in Tb also gained productivity in the post-test. Results from comparing inter-group productivity gain in Tb showed a medium effect size which indicate that MTPE cannot be ignored as a method to increase productivity in translation even though the productivity gain for students in Gb was not statistically significant in comparison with students in Ga when total translation times in both time intervals were taken into consideration.

As for the first two parts of the productivity analysis, i.e., the longitudinal analysis of the productivity gain for both groups (sections 4.3.1 and 4.3.2), there was a significant productivity gain for both groups in Tb when compared to Ta with Gb recording larger productivity percentage than Ga (on a within-group level). A possible explanation for the evident productivity gain in both groups is that students have gained productivity in Tb because it is the expected outcome as a

result of four weeks of translation training in class. This result goes in line with Colombo and Stanca (2014) as well as Cherif (2021) who found that training has a positive and significant effect on productivity. Thus, the factor of more training leads to more productivity was validated for both groups. A possible explanation for the larger productivity gain (33%) in MTPE when compared to that of students who used HT (19%) in the current study is that students who HT, being inexperienced translators, tend to deal with the translation task as a lexical task (Tirkkonen-Condit 1990) which is validated by the students' overreliance on dictionary use during the translation task (Alsalem 2019). In Jia et al. (2019), non-professional translation students who also considered translating from English to Chinese a lexical task, showed similar increase in productivity when they used MTPE compared to HT. In addition, the results from the current study align with those from Aranberri et al. (2014) in the sense that although the participants did not get a chance to read the contents of the texts assigned for the translation tasks prior to Ta or Tb, their familiarity with the topics (clogged showerheads and instructions to operate a washing machine) was in fact beneficial and reflected positively on their increased productivity when post-editing MT output.

Although it does not support the larger percentages reported in previous studies, the within-group percentage of productivity gain (33%) in Gb cannot be ignored. There are a few possible explanations for the relatively smaller productivity gain of students in Gb in the current study when compared to productivity gain from previous studies (e.g., Haji Sismat 2016; Plitt and Masselot 2010): one might be due to the lack of experience amongst students being undergraduate with no professional experience in translation, and two, is the relatively short period of time allocated for the intervention. An additional possible explanation for this result may be linked to students' perceived opinions of MT-related skills that were previously expressed in the FGDs, i.e., teachers considering the use of MT in the classroom as cheating and warning students against using it (Al-Jarf 2017). In addition, the recurring errors in the Arabic MT output might have caused the students to spend more time trying to correct them. Finally, a possible explanation can be drawn from the results of Al-Gahtani et al. (2007) in which women showed more reluctance towards technology and less acceptance of it. As the current study is single-sexed, it could be that female students were more reluctant to accept MT suggestions which might have affected their productivity.

What has emerged from the inter-group productivity gain comparison (section 4.3.3) is that the productivity gain in Tb for students in Gb just fell short of being statistically larger than that of students in Ga. Again, this result may have to do with the relatively short training period as students in Gb have had 4 weeks of MTPE training while students in Ga have been training to use HT and work under time pressure for the past five semesters prior to the level in which they

participated in the experiment. However, when considering the increase in the average productivity gain in Gb (28%) independently, it clearly shows that MTPE has helped improve students' productivity through post-editing. This result corroborates that of Haji Sismat (2016) who found that MTPE in English-Arabic translation for Malay students who are non-native trainees of Arabic was 14 times faster than HT. Furthermore, it supports Green et al. (2013) who found that PE leads to reduced time for professional translators translating in the language pair English-Arabic. Although the percentage is lower, this result also supports Plitt and Masselot's (2010) finding that MTPE increased the productivity by 43% for professional translators

When referring back to the findings from the FGDs, 75% of students thought that MTPE was faster than HT. A percentage which was validated by the performance of students in Gb after the intervention. However, when asked about their opinions in the retrospective pre-test survey, some students in Gb thought that HT was faster although they have shown significant productivity when they used MTPE. A possible explanation for this is due to the more cognitive or technical effort that they had to endure while post-editing MT output. Lee and Liao (2011) note that students expressed that MT output was slowing them down as they were thinking of better ways to refine the TTs. Also, Doherty and Moorkens suggest that the perception of MTPE amongst students could either be a 'learnt attitude' (2013: 132) or that biased perception could be resulting from students not having enough technical background to handle MTPE tasks easily. Thus, it would be beneficial to include hands-on training on the technical aspects of MTPE if it is to be introduced in the undergraduate classroom of students coming from backgrounds with little or no technical expertise. Such training which aims at increasing the technical competence may include but is not limited to (1) training on handling different MT systems (e.g., Google Translate, Amazon Translate, and Trados), (2) training on different CAT tools (e.g., translation memories, language search-engine software, and interactive MT), and (3) increasing error-prediction amongst students which can be achieved through practice to lower keyboard effort. In addition, an in-depth explanation provided by the teacher of the course about the cognitive processes during an MTPE task might assist the students when they take on a task that is not as straightforward as translating from scratch. Further research on the cognitive and technical efforts in English to Arabic translation in the classroom is thus recommended.

The next section discusses the results of translation quality.

## 5.4 Discussion of Translation Quality

This section discusses the error type and error count of MTPE compared to HT in an attempt to answer **RQ3:** What are the differences in the product between HT and MTPE?

- What are the most common errors in HT and MTPE tasks in the language pair English-Arabic?
- Is there a difference in the overall quality between the product of HT and the product of MTPE output?

The aim is to evaluate the effectiveness of MTPE through examining the quality of the TTs resulting from MTPE compared with HT. It seeks to answer RQ3 about the differences in the product between HT and MTPE by examines the HT and MTPE of the same group (Gb) to answer the first sub research question of RQ3 about the most common errors in HT and MTPE tasks in the language pair English-Arabic. It also explores whether the students in the MTPE group (Gb) score similar or better results when compared to students in the HT group (Ga) to answer the second sub research question of RQ3.

The discussion section begins with a summary of the results generated from the analysis of both error type and error count. Following are sections allocated to elaborate on the discussion of the results in the light of related studies.

The analysis of error types has identified that while MTPE helped students avoid deletion and technical errors, accuracy and comprehension as well as grammatical errors increased in MTPE translated texts. On the other hand, error count results (section 4.4.2 above) broadly support prior studies which compared MTPE with HT (e.g., Haji Sismat 2016; Plitt and Masselot 2010; Yang et al. 2020) which suggest that if MTPE produced similar or better quality when compared to HT, then it can be concluded that it is effective. In the current study, Ga students who used HT in Ta and Tb gained better translation scores in Tb, Gb students who used HT in Ta and MTPE in Tb also gained better translation scores in Tb, and results from the inter-group comparison of translation quality scores showed a zero-effect size which indicates that students who used MTPE have increased scores in a similar manner of those who used HT but not better.

### 5.4.1 Error Type

It was mentioned earlier that the error type analysis was carried out to identify the most common errors in HT and MTPE so that these error types are highlighted in PE training and to Arabic MT developers. This analysis was run on the TTs generated by the students in Gb in both Ta and Tb. The results showed that out of four error categories assigned for the analysis, MTPE helped improve deletion and technical-aspects errors. On the other hand, MTPE seems to have increased the errors related to comprehension, accuracy and register in addition to errors in grammar, cohesion, coherence and organization of work. When considering the overall results, we can generally say that they do not corroborate those of Jia et al. (2019) in which MTPE generated

equivalent TTs in terms of accuracy and fluency. Thus, scrutinizing the results would prove beneficial when the aim is to evaluate the effectiveness of MTPE in the classroom.

When examining the results from the current study in light of those from Al Mahasees (2020), we notice that the PE training intervention managed to help students in Gb improve some errors that were previously mentioned in Al Mahasees' study. Namely, PE has decreased Al Mahasees' previously identified errors in capitalization, punctuation, numbering, and agreement in MT output. However, the previously identified errors of mistranslation and wrong word order do not seem to have benefited from the post-editing of MT output. Therefore, the results from the current study confirm Al Mahasees' expectation that PE would be necessary to generate acceptable final translations. Yet, the current results suggest that either MT developers focus on improving these shortcomings in the quality of Arabic MT output, or PE training is essential and should focus more on fixing errors related to comprehension and grammar in Arabic MT output (Yamada 2019).

The results of the current study which revealed that MTPE has improved students' performance in relation to deletion and technical aspects do not confirm those of Izwaini (2006) in which deletion and spelling mistakes were found to be major problems when using the same Arabic MT system that is used in the current study, i.e., Google Translate. This contradiction can be attributed to the recent improvements on the Arabic MT output especially as Izwaini's study was conducted more than a decade ago using statistical machine translation (SMT) whereas the current study used the improved NMT service provided by Google Translate. However, the results from the current error type analysis confirm Izwaini's results regarding identifying errors in lexis, grammar, and style which might indicate that the current NMT still cannot avoid the different errors in Arabic MT output caused by lexical and grammatical differences between English and Arabic. In addition, due to the fact that English and Arabic belong to different language families (Al-khresheh and Almaaytah 2018), the current results which reveal more grammatical errors in the Arabic MTPE text confirm Depraetere's (2010) claim that post-edited texts tend to stay closer to the structure of the source text. This can be clearly noted in the reviewers' annotations of the MTPE texts where they frequently wrote 'awkward translation' and 'translation sounds awkward' when commenting on TTs that were post-edited even though they did not know the method of translation used for each translated text.

Although the current study used a NMT system, the results from error type analysis confirm results from Haji Sismat (2016) which used a SMT system and revealed that lexical and syntactic errors appeared more in MTPE texts than in HT texts. They also confirm Al-khresheh and Almaaytah's (2018) result that the questionable accuracy of Arabic MT output still exists. The

current study would affirm the recommendation of the latter authors which stated that full PE was essential especially to maintain the accuracy and quality of Arabic MT output. Thus, we recommend that in order for translation students to make full use of MT output, that PE training is integrated into undergraduate translation programmes to prepare students for the job market.

Inconsistency of terms was also an issue that the reviewers noted in the MTPE translated texts. This result corroborates with that of Guerberof (2009) who found that HT was better regarding consistency. A good example can be extracted from the current study in which an English proper name was found in the text used for Tb (the word Original with a capital O). Students in Ga who translated the same text using HT knew that it was a proper name and not an adjective because of the capitalized (O), therefore they transliterated it into the Arabic (أوريجينال – *əˈrijənl*). However, the MT system did not realise the difference between the adjective (original) and the noun (Original) and therefore it translated it into the meaning of (original) in Arabic (أصلي- *ʻasli*).

All in all, the results from error type analysis confirmed that MTPE helped improve errors in deletion, capitalization, punctuation, numbering, and agreement, however, errors in comprehension and grammar still existed. These types of errors suggest that the Arabic MT output is quite poor when attempting to maintain the correct transfer of information and evidence of complete comprehension and appropriateness of rendering and lexis.

### 5.4.2 Error Count

The ultimate purpose of the error count analysis was to evaluate the effect of MTPE on the translation scores generated by translation students in the classroom. In order to do so, three types of analysis were run, i.e., horizontal analysis of Ta for both Ga and Gb, overall horizontal analysis of both Ga and Gb, and lastly, a longitudinal comparison of Tb against Ta of Gb.

The horizontal error analysis in Ta was carried out to determine the translation abilities of students in Ga and Gb and to calculate the baseline scores against which the later comparison is to be conducted. The results showed a slight disparity between Ga (78 points) and Gb (74 points) in Ta which indicates that students in Ga have generally performed better than those in Gb.

The horizontal between-group error analysis of Tb was carried out to evaluate the effect of MTPE on Gb when compared to students in Ga who used HT. When the effectiveness of MTPE is evaluated based on score change, the result showed that students in Gb have achieved quite similar score difference when compared with students in Ga as the total score increase of Gb in Tb was 6 points whereas the score increase of Ga was 5 points. These results confirm the results from Guerberof (2009), Daems (2016), and Jia et al. (2019) which all revealed that MTPE did not

grant significantly better quality when compared to HT in terms of scores. A possible explanation for this result in the current study can be extracted from the error type analysis that was run on the MTPE group which revealed that although MTPE has helped students avoid numerous errors caused by deletion and technical errors, however, errors related to accuracy, register, and grammar have been mentioned more when students in Gb used MTPE than when they used HT (further discussion of error type results is in section 5.4.1). Even though the current evaluation of Arabic MT output (Google Translate) has shown lexical and syntactical errors which greatly affected the quality of the MT output causing it to become unintelligible (Jabak 2019), and regardless of the relatively short period of MTPE training, Gb students who used MTPE have managed to gain scores in Tb in a similar fashion to those in Ga which again supports Garcia's (2011) suggestion that students can benefit from PE training if they show good performance in MTPE tasks without prior experience.

However, the current results from between-group comparison do not support those by Yang et al. (2020) and Garcia (2011) which concluded that students who use MTPE performed better than those who used HT regardless of the language pair used, the level of experience or the translators' level of performance. When I returned to both studies, I noticed that no PE training periods were assigned for the participants in their studies. Therefore, a possible explanation that remains is the improved quality of Chinese MT output as both studies were conducted using the language pairs English-Chinese-English. This suggests that if Arabic MT output is improved, it may result in better MTPE performance by inexperienced translation students, and more productivity for the experienced ones as they will not spend more time on fixing reoccurring errors.

On the other hand, the longitudinal within-group error analysis of Gb was carried out through measuring the change in scores of TTs of the same group of students using HT in Ta and MTPE in Tb (i.e., Gb scores). The result revealed that the average mean score of Gb has increased by 6 points in Tb than in Ta, which is a result that we discussed in the previous section. However, an interesting pattern appeared when students' scores were represented through a graph (see Figure 4-14). The graph showed a pattern in which the worst-scoring students in Ta (HT) seem to have made the best use of MTPE in Tb (evident in their positive score difference), whereas the best-scoring students in Ta performed worse when they used MTPE (evident in their negative score difference). This pattern affirms the conclusion reached by Varela-Salinas (2020) which states that CAT tools can compensate to a certain extent in the case of poor linguistic knowledge. However, the pattern of the better-performing students in Ta did not confirm the second part of Varela-Salinas's conclusion which states that students who master a certain level in both the SL and the TL were the ones who benefited the most from exploiting technology. This pattern may again confirm that the quality of the Arabic MT output is low in terms of lexis and structure

because students who did not apply all PE skills in Tb lost points due to errors in comprehension, accuracy, and grammar whereas students who gained better scores in Tb did so because they avoided deletion and technical errors that their results in Ta revealed. In addition, if we are to consider students in Gb who scored the least in Ta to broadly share similar profiles with lay-users, their results may corroborate the results of Aranberri et al. (2014) which pointed out that lay-users were the most to benefit from MT output. The current results suggest that less-experienced translators and lay-users are the ones to benefit the most from the current quality of Arabic MT output. Also, the fact that the better-skilled translation students did not do well when they used MTPE affirms that MTPE requires skills that differ from those required for translation from scratch, and that Arabic MT is still in serious need of revision and development.

To try to locate the aspect of performance on which MTPE was specifically effective, I examined the results and the survey responses of the 9 students whose scores increased by 10 points or more. The criteria behind the decision of examining only these results stems from the desire to thoroughly examine the range of significant score differences and not include all in an attempt to find a pattern. Out of 31 students (Figure 4-14), the scores of 9 students have increased by 10 points or more. When I examined their mark sheets, I noticed that their scores generally increased due to better performance in AoP2 (grammar, coherence, cohesion and organization of work) and AoP3 (technical aspects) (See Table 5-1 below)

Table 5-1 Score differences in all Aspects of Performance for the best scoring students in Gb

| Student Chart ID | Ta Score | Tb Score | Score Gain | Ta-AoP1 (50) | Tb-AoP1 (50) | Ta-AoP2 (35) | Tb-AoP2 (35) | Ta-AoP3 (15) | Tb-AoP3 (15) |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 49 | 75 | 26 | **24** | **41** | **19** | **25** | 6 | 9 |
| **2** | 55 | 73 | 18 | 37 | 29 | **13** | **30** | **5** | **14** |
| **4** | 61 | 85 | 17 | **31** | **39** | **21** | **32** | **9** | **14** |
| **5** | 64 | 82 | 18 | 39 | 37 | **18** | **34** | **7** | **11** |
| **6** | 64 | 77 | 13 | 39 | 36 | **16** | **28** | **9** | **13** |
| **7** | 68 | 84 | 16 | 35 | 34 | **21** | **35** | **12** | **15** |
| **11** | 72 | 85 | 13 | 42 | 41 | **22** | **33** | **8** | **11** |
| **13** | 72 | 86 | 14 | 41 | 40 | **21** | **34** | **10** | **12** |
| **22** | 79 | 96 | 17 | **41** | **46** | **28** | **35** | **10** | **15** |

With closer examination of Table 5-1 above, I noticed that all Gb students who scored significantly better in Tb have performed better in AoP2 (scores that have increased in Tb are underlined in

Table 5-1 above) which indicates that MTPE may have enhanced the performance of translators in AoP2. However, this enhancement in AoP2 performance cannot be attributed to PE skills that students have learned in the course. Actually, the comparison between the number of AoP2 errors found in raw MT output (6 errors) and those found in MTPE (13 errors) indicating that students' overall tendency to overcorrect contributed to their increased errors in AoP2 (figure 4-12 above). I say this because some of the most frequently recurring errors that the evaluators annotated in the error type analysis (section 5.4.1 above) were of grammatical nature. In addition, eight out of the nine students whose scores have increased in Tb performed better in AoP3 (technical aspects) while only 3 students have gained more scores in AoP1 (comprehension, accuracy, and register) which indicate that MTPE has helped students avoid technical errors but not errors related to comprehension and accuracy. This again corroborates the conclusion of Jabak (2019) which suggests that lexical and syntactical errors greatly affect Arabic MT output. This result supports the call for Arabic MT developers to focus on fixing these comprehension-and-accuracy-related errors that even NMT systems could not avoid. In addition, it highlights a need to focus on fixing these types of errors in PE training.

The close examination of the TTs of the students whose scores have significantly increased in addition to their responses in the after-survey also suggests that Davis' (1989) 'acceptance' (which entails finding technology both useful and easy to use) exists. All 9 students revealed in their after-survey responses that they 'find the output of MT systems useful' and 'easy to handle'. This result highlights the importance of acceptance when promoting technology use amongst students and supports the conclusions provided by many researchers such as Jia et al. (2019), Daems (2016), and Yang et al. (2020).

## 5.5 Chapter Summary

In this chapter, I have discussed the results reported in chapter 4. The discussion tackled RQ1 about students' opinions for which two instruments were used. Students who participated in the FGDs generally showed reluctance and scepticism towards MT and PE. However, students who participated in the PE teaching intervention showed the tendency to embrace MTPE training as they thought it comprised an essential part of the work environment. In addition, this chapter discussed results of RQ2 about translation productivity which indicate the productivity gain from MTPE as statistically medium in terms of effect when compared to HT. The importance of the effect size here stems from that fact that when the results are statistically non-significant (such as in the resulting non-significant p-value in the current study), the difference in productivity cannot

be neglected. In other words, this medium effect size indicates that if the sample size was larger, that the p-value would be statistically significant. Finally, this chapter presented the discussion of error count and error type to answer RQ3 about translation quality. Results from error count showed that MTPE-students managed to gain similar scores to HT-students but not more. Examining error types highlighted the reoccurring recommendation to focus on fixing errors related to accuracy and grammar whether it be through PE skills or MT development.

# Chapter 6 Conclusions

## 6.1 Introduction

This chapter summarises the main aspects covered in the current study. It begins with the thesis overview (section 6.2) that lists the main RQs, findings from the analysis of data gathered to answer those RQs. Then it addresses the research questions in section 6.3 and provides the contributions of the study in section 6.4. Limitations and difficulties experienced during its execution are listed in section 6.5. Finally, this chapter ends with a list of recommendations and suggestions for future research in section 6.6.

## 6.2 Thesis Overview

This thesis offers a mixed-method experimental approach to evaluate the effectiveness of MTPE in an undergraduate translation programme in the language pair English-Arabic to answer the following three research questions:

**RQ1:** What are the differences in students' opinions about HT and MTPE?

**RQ2:** What are the differences in the process between HT and MTPE?

**RQ3:** What are the differences in the product between HT and MTPE?

In the literature review, a research gap was identified as the absence of a formative evaluation of the effectiveness of MTPE in undergraduate translation programmes that tackle the English-Arabic language pair. This gap in literature was identified as a result of delving into three areas in research: studies on ways to bridge the gap between translation training and the translation industry, studies about errors in Arabic MT systems, and studies which evaluated Arabic MT. The gap still existed when I commenced the study although a number of studies recommended closing the gap between translation training and the translation market through proper exploitation of CAT tools and CAT training. In addition to my personal interest in the subject, the importance of evaluating the means to bridge this gap was highlighted through studies that reached the conclusion that Arabic NMT systems still generate output that requires fixing. However, studies evaluating the efforts, tasks, and processes involved in MTPE have rarely touched upon English-Arabic, and they did so in contexts that differ from the one tackled in the current study.

In response to the above research needs, a conceptual framework was chosen (Kirkpatrick Model of Learning Evaluation) to run the formative evaluation of MTPE in the translation classroom. Chapter 3 was devoted to the methodology. It provided detailed discussion of the philosophical

stance behind the current study, justifications for the mixed-method approach used, research settings. In further sections, the chapter provided detailed discussion of the sampling criteria, TQA criteria and details about the teaching intervention that was given to the students. The chapter also covered methods of data collection and data analysis used in every step of the Kirkpatrick Model of Learning Evaluation, in addition to the ethical considerations.
The results (chapter 4) and the discussion (chapter 5) of using a mixed-method approach were presented in an order that follows the order of the RQs as well as the Levels of the Kirkpatrick Model. These chapters further extended the discussion through linking the current results with results and conclusions of previous literature.

Chapter 4 presented the findings and results of the data derived from FGDs, retrospective pre-test surveys (instruments used to answer RQ1), translation task time (RQ2), errors in translation and translation scores (instruments used to answer RQ3). To answer RQ1 about students' opinions (Level 1 on Kirkpatrick Model), this chapter provided thematic analysis of transcriptions derived from recorded discussions that were held before the intervention of 26 students. In addition, descriptive analysis of survey responses that were collected after the intervention were used to triangulate the results of the first set of data. One of the major findings was that although students showed tendency to present biased opinions about MTPE degree of reward before the intervention, their opinions became more balanced and fairly positive as shown in the post-survey after they have learnt about MTPE training and use. Students were also divided about the usefulness of MTPE but showed a sharp increase in their opinion about it being useful after they had some hands-on MTPE training and practice. Also, comparing the before and after survey responses about MTPE ease of use showed that the intervention was successful in promoting that MTPE was user-friendly amongst students. Another significant finding was that, after the intervention, the number of students who thought that HT is faster has increased indicating that some students found that MTPE required either more technical or cognitive effort although the findings from the FGDs showed that students thought MTPE was faster than HT, and despite the responses that indicated that MTPE was user-friendly after the intervention. This part of the study also investigated students' opinions about the translation quality of both HT and MTPE. Most students were biased and thought that the quality of HT was better. This view has changed in the post-intervention survey where the majority of students thought that the quality of MTPE was as good as HT or better. The positive change towards MTPE in students' attitudes was obvious in the final question in which they were asked about their preferred method of translation. Although the responses from the FGDs and the pre-test survey showed that the majority of students preferred HT, a steep increase in the number of students who preferred MTPE has been noticed in the post-intervention survey.

Chapter 4 also presented statistical results to answer RQ2 about comparing translation productivity between MTPE and HT. I have investigated the productivity of MTPE through comparing group total task times. Students in both groups managed to gain more productivity in Tb, but students in Gb did not present productivity gains that reflect results from previous literature although their results were promising. Therefore, the overall results answered RQ2 suggesting that MTPE helped improve students' productivity but that more awareness of the technical and cognitive efforts to increase the productivity of the task are recommended.

Also, chapter 4 detailed statistical results drawn from error type and error count analysis to answer RQ3 about translation quality. The analysis of error types has identified that although MTPE helped students avoid deletion and technical errors, nevertheless, accuracy and comprehension as well as grammatical errors appeared more in MTPE translated texts. On the other hand, error count results broadly support prior studies which compared MTPE with HT (e.g., Haji Sismat 2016; Plitt and Masselot 2010; Yang et al. 2020), and which suggest that if MTPE produced similar or better quality when compared to HT, then it can be concluded that it is effective. In the current study, Ga students who used HT in Ta and Tb gained better translation scores in Tb, Gb students who used HT in Ta and MTPE in Tb also gained better translation scores in Tb, and results from Gb inter-group comparison of translation quality scores showed a zero-effect size which indicates that students who used MTPE have increased scores in a similar manner of those who used HT but not better.

Chapter 5 discussed the findings and results of the current study in light of previous literature. It began by discussing results from gathered data about students' opinions. For instance, students' reluctance about the productivity of MTPE after the completion of the teaching intervention was attributed to possible technical or cognitive efforts that the students had to endure during the MTPE task, and therefore recommendations for running further investigations on technical and cognitive efforts in MTPE tasks were provided. As for the actual productivity of students in the current study, chapter 5 discussed possible reasons for why the students did not reveal significant productivity gain when they used MTPE such as the allocated time for the teaching intervention and therefore allocating more time for practice was recommended. Finally, regarding translation quality, the discussion of error type suggested that two aspects of performance may benefit the most from focused MTPE training. Namely, the (1) comprehension, accuracy and register, and (2) grammar, cohesion, coherence and organization of work. In addition, the discussion of error count revealed that although students who used MTPE managed to achieve similar score gain to those who used HT, that focused MTPE training and development of Arabic MT output may increase the quality of MTPE TTs amongst students of English-Arabic translation.

## 6.3 Addressing the Research Questions

This section provides answers to the research questions detailed in the first chapter (section 1.3).

**RQ1:** What are the differences in students' opinions about HT and MTPE?

When referring back to the aims of implementing the Kirkpatrick model of learning evaluation, evaluating the reactions, or students' opinions in the current study can be achieved through attempting to answer the questions: How do students feel about MTPE? Do they find it useful? Therefore, if the present study found that students' opinions are or have become positive about the usefulness of MTPE and found it to be fairly easy to use, MTPE would prove effective in fulfilling the goal behind evaluating the first level.

In order to determine whether the expectations of MTPE training and use have been fulfilled in terms of students' opinions (i.e., prove effective), the following observations have been made: Overall, students' responses suggest an appreciation of the importance of being able to learn and use MTPE. The view was expressed by many students who thought that incorporating MTPE reflects an image of a workplace that they would like to embrace.

While reluctance and scepticism have comprised a major part of the opinions of the first group of students in the FGDs (the group that did not participate in taking the teaching intervention), the tendency among the second group of students (who took the intervention) to accept MTPE after they have tried it provides a significant rationale for curriculum innovation that encourages MTPE training and use within the classroom. After the intervention, responses regarding the degree of reward, the usefulness, the ease of use, the productivity and the quality of translation tended to be positive. This indicates high satisfaction levels with MTPE as a method of translation. Indeed, this was reflected in the responses of the students who took the intervention to the last question in the survey about their preferred method of translation as most of them shifted their preference from HT before the intervention to MTPE after.

Nevertheless, opinions about translation productivity amongst students who participated in taking the intervention did not show a significant difference in the after-survey. According to Krings (2001), PE effort is classified into three distinct categories: temporal (the time spent), technical (the physical action), and cognitive (the mental processing). In Krings' view (2001), temporal effort is the result of combining both the technical and the cognitive efforts. When applying Krings' explanation, it is possible that the students who found PE to require more temporal effort expressed this view because of the technical effort, or the work involved in the keyboarding and mouse actions needed to fix the errors in the MT output, or because of the cognitive effort, or the mental processing involved. It is, thus, useful to consider researching PE

efforts in the English to Arabic translation in the classroom to try to find out an explanation for the students' opinions especially that they have actually showed significant MTPE productivity results in the post-intervention translation task (see section 4.3.2).

**RQ2:** What are the differences in the process between HT and MTPE?

According to the aims of implementing the Kirkpatrick model of learning evaluation, evaluating the process, or in this case, the translation productivity can be achieved through attempting to answer the question: is the skill that the course aimed to achieve present? Therefore, if the present study answered the research question through demonstrating productivity gain amongst students who used MTPE, MTPE would prove effective in fulfilling the goal behind evaluating the third level.

The study employed a pre-post measurement and comparison of total translation time in minutes to allow for accurate statistical analysis of the performance of students in order to decide whether MTPE was faster than HT. The result from this part of the study leads to a tentative conclusion that the productivity gain from MTPE cannot be ignored as it showed a statistically medium effect when compared to HT. However, due to the different tasks employed in the two groups, conclusive results of the productivity of MTPE in comparison to HT cannot be made (see section 6.4). Nevertheless, on a within-group analysis level, students' overall productivity gain was 33% which suggests that the integration of MTPE in translation training is most likely effective. Garcia (2011) suggests that if students without prior MTPE experience show good performance in MTPE, then their performance can benefit from training in PE skills.

**RQ3:** What are the differences in the product between HT and MTPE?

According to the aims of implementing the Kirkpatrick model of learning evaluation, evaluating the product, or in this case, the translation quality can be achieved through attempting to answer the question: What is the impact of MTPE on translation quality? Therefore, and based on previous studies, if MTPE guaranteed similar or better quality to that of HT, then it would prove effective in fulfilling the goal behind evaluating the fourth level.

In-depth analysis of the HT and MTPE error types of the same group of students revealed that two major AoPs may benefit the most from focused PE training in the classroom, and from accurate improvement of the Arabic MT output carried out by MT developers. These two AoPs are namely the (1) comprehension, accuracy and register, and (2) grammar, cohesion, coherence and organization of work. In addition, results from error count led to a tentative conclusion that even though MTPE managed to reduce the time required to complete the task by 33%, it generally did

not assist the students to score better in terms of translation quality. However, in terms of translation quality, students who used MTPE managed to gain similar score increase to those who used HT although MTPE has produced more errors related to accuracy and grammar. This result is portrayed in the score change of students in the MTPE group that increased in a similar fashion to students who used HT. However, due to the different nature of translation tasks implemented in the two groups, a definite conclusion on whether students who used MTPE performed better or worse than students who used HT cannot be made in the current study (see section 6.5).

According to the results of students in the MTPE group, MT output in its current shape is likely effective in the translation training of undergraduate translation students when translation quality is the purpose of the practice, but further improvements on MT systems and guidelines of PE training are required if more effectiveness is sought.

## 6.4 Contributions of the Study

This study is the first to explore the opinions of undergraduate all-female translation students while simultaneously attempting to explore the effectiveness of MTPE training on translation productivity and quality in a holistic systematic formative evaluation of the language pair English-Arabic. This thesis took a focused MTPE training course that was derived from TAUS PE guidelines as a sample of the specialized translation courses in an Arab translation college and evaluated its effectiveness in bringing about more positive opinions and faster translation without compromising the quality. Through its findings, it is hoped that it will be the first study to offer guidelines not only on how to encourage undergraduate students to accept MTPE training, but also recommendations on how this type of training can be incorporated into the undergraduate translation curricula as a step to bridge the gap between translation training and the translation industry. The following section provides a list of guidelines stemming from the findings of the current research. After that, a summary of the empirical, methodological, and theoretical contributions is provided.

### 6.4.1 Guidelines for Arabic MTPE Stakeholders

In light of the findings and discussion of the current study, this is a proposal for MTPE guidelines in translation programmes that tackles the language pair English-Arabic. It is important to highlight that these recommendations are drawn from an exploratory study that was conducted in one college with one sample of students who used one MT system (Google NMT). These guidelines would always benefit from further research on their validity in similar academic programmes and in larger-scale studies.

**Guidelines stemming from students' opinions**

1. Students should be made aware of the potentials of MTPE while, rather than advising them against exploiting it, bringing their attention to the notion that it is not technology that risks replacing translators in the industry, but a technology-oriented translator. In other words, as in every other field of knowledge, students should know that technology is dedicated to complement a translator's job and not replace it.
2. Browsing job advertisements that require technical expertise with a focus on MTPE skills at the beginning of a PE course would enhance the sense of usefulness amongst students. This simple purposeful action may provide students with the required motivation to learn and practice PE skills.
3. Students' expectations should be made realistic. They should be made aware of the fact that due to the drastic linguistic differences between English and Arabic, neither Arabic MT output alone nor in conjunction with PE are the way forward with all text types (such as literary translation which still requires further investigations, and the translation of the meanings of the Holy Quran). Simultaneously, students should be made aware of the potentials of MTPE in gaining productivity while maintaining human-translation quality for specific text types, such as technical texts and reports.
4. As for students who think that HT is easier than MTPE, practicing PE skills may be the way forward. This is being said because error prediction, a skill that is gained through practice, is considered one of the skills that increase productivity and a factor in finding PE process free from cognitive effort.

**Guidelines stemming from productivity results and discussion**

5. Students should be made aware of the levels of PE, i.e., light PE and full PE, and that even full MTPE entails gaining productivity without compromising the quality.
6. Practicing different MT systems (such as Google MT, Amazon MT, and Trados) is the way forward for mastering error prediction. This skill in turn increases productivity as the student would gain productivity through purposeful skipping and or pausing to fix previously predicted errors of each particular MT system.
7. Focusing on skills that lower both cognitive and technical efforts. Such skills may include hands-on technical training that includes keyboard skills and shortcuts that would decrease the technical effort associated with MTPE.
8. Demonstrating that not every error requires fixing
9. Allowing students to examine the differences between TTs produced through MTPE and HT.

**Guidelines stemming from translation quality results**

10. Focusing the training to exploit the benefits of the current Arabic NMT. That includes its ability to avoid errors in deletion and technical errors.
11. Including training that focuses on avoiding current errors in Arabic NMT such as errors in accuracy, comprehension and grammar.
12. Highlighting that PE requires skills that are different from those required for HT, teachers could assist students who perform better in HT but not as well in MTPE through exercises in which students play the roles of MT developers. In such exercises, students focus on revising MT output suggesting solutions for the errors they spot.

**Guidelines for educational stakeholders (program directors, teachers, MT developers)**

13. Stakeholders should aim for designing PE curricula and updating it based on adaptation of the market needs. That is, designing PE curricula shall consider training on the most widely used MT systems in the market and training on PE skills of the mostly demanded text types.
14. Even with the improved Arabic NMT, results from MTPE tasks in Arabic show that errors in comprehension, accuracy and register as well as errors in grammar, coherence, and cohesion still exist and exceed similar error types in HT. Based on this result, it is recommended that MT developers pay closer attention to the contexts of Arabic MT output and link that to the proper choices of word. Meanwhile, and until further development is noticed in Arabic MT output, teachers should focus students' attention on these types of errors when they are being trained to post-edit MT output as these errors tend to occur in higher percentages.

### 6.4.2 Empirical, Methodological, and Theoretical Contributions

Empirically, it is hoped that the guidelines listed in the previous section (section 6.4.1) will reflect on how MTPE training courses are tackled by stakeholders in order to ensure that students are accepting technology-use to achieve productivity in translation without compromising the quality of the TTs. Surely, as the findings proved that the training course was successful, the training course could be made more widely available possibly after being revised and updated in light of the findings from this study and the accumulated output through delivering it.

As for the methodology of this thesis, there are several studies which explored the opinions of translation students, several studies conducted on the productivity and quality of translation

through the use of MTPE, and some studies that examined the three aspects. However, this is the first study that illustrates a methodology which evaluates all three aspects in one study that focuses on the English-Arabic language pair: Students' opinions before and after the teaching intervention, and attempting to link the findings from exploring students' opinions to results from an experiment in which the productivity and quality of TTs produced by the experimental group (students who used MTPE) are compared with those of the control group (students who used HT) in a holistic evaluation of learning. It is also hoped that the design of this research is replicated in other translation programmes and in other language pairs to compare results, and to illustrate how classroom studies of undergraduate students' interaction with MT and their opinions about it can be conducted.

Theoretically, the thesis hopes to add towards literature on translation studies. Namely, *translation criticism* (studies that focus on the evaluation of students' performance in translator training) as well as literature on *translation aids* (i.e., MT strategies), and finally, literature on *teaching methods* within translator training.

## 6.5 Limitations and Difficulties Experienced During Research

This study utilized a rigorous mixed-method approach and an experimental design to evaluate English-Arabic MTPE training in translation undergraduate programs. Although this approach provided valuable insights and perspectives, some limitations and difficulties need to be acknowledged.

Firstly, the methodology of this study did not come without challenges. For instance, even though I tried to take as many factors as possible into account in my data collection, I could not control everything. For example, when conducting the FGDs, the original design was balanced seeking to recruit five to six students per discussion so that the discussion is richer as per recommendations from previous literature; each student is given more time to participate, and more data can be collected. However, due to some students' desire to postpone their participation in the discussions to a later date because of mid-term exams, and the fact that I was committed to a flight back to the UK, two discussion groups had to be merged and the last FGD had a total of twelve students. To mitigate against this imbalance, I sought the help of a colleague in order to manage the last discussion and to make sure each student was given enough chance to participate. Also, rather than allocating one hour for the discussion like the previous groups of four or five students, two hours were allocated for this larger group in order to give each participant a chance to speak her mind.

In addition, although the data collection plan intended to have every student who participated in the experimental group to fill out a survey after the translation task, not all students agreed to participate. Four students decided to leave the language lab once they finished the translation task and they apologized for not taking part in filling out the surveys. As in every other research, more collected data means more robust findings. Thus, it would be interesting to compare the findings of the current study with large-scale studies.

Also, the sampling technique used when recruiting for this study (cluster sampling) had yielded baseline disparities that affected students' results as an extraneous variable, and it resulted in time-management challenges. Although this sampling technique was recommended in pedagogical studies of translation (Saldanha and O'Brien 2014), nevertheless, when comparing productivity, the natural cluster of students in Gb scored better than those in Ga in Ta which made it impossible to validate the effectiveness of MTPE on productivity without changing the statistical tool used for analysis (I used *t*-tests the first time then switched to ANOVA the second time). Although students in Gb started off in the pre-test (Ta) by showing slightly shorter task times than students in Ga which one might think is a normal cause for the higher productivity gain in Tb, however, the proportional productivity in Tb was higher for Gb than it was for Ga (15 minutes to 10 minutes respectively). This yet confirmed the need to use a statistical tool that took all factors into consideration at once, which added an unnecessary challenge of learning and executing ANOVA at a later stage of research. Although previous research found 'hardly any differences between estimates, CI's and p-values from adjusted and unadjusted analyses' (De Boer et al. 2015: 4), it would be interesting to compare the results of the current study with one that has adjusted randomization to ensure similar baseline results.

Furthermore, task duration was calculated subjectively (similar to Lee and Liao 2011). The intended use of accurate time applications and software was not possible at the time of data collection due to two reasons: the computers in the labs had technical issues that prevented downloading and installing software such as Translog-II or Time Doctor while the technical support in language labs was scarce. Because not all students had a personal laptop, it was inevitable to use the computers provided in the labs and calculating translation task times subjectively.

In addition, due to the fact that the two cohorts did not complete the same task where one group practices PE without taking the intervention and the other practices it after taking the intervention, it is difficult to assert the effectiveness of the MTPE intervention in this study.

In terms of results, it may not be possible to generalize the findings of the study to a further population than those similar to the one addressed in the current study. The obtained results are based on a homogeneous small sample of participants. As for homogeneity, the translation students were all females from a public university in the central region of Saudi Arabia. However, the influence of gender on MTPE tasks and processes could not be investigated as it was not an intended decision but rather the natural representation of college students that was easier for my access. Finally, results of this study are based on the performance of novice translators. A similar study conducted on experienced translators would have provided more robust results.

## 6.6 Recommendations for Future Research

### 6.6.1 Allocated Time for the Study

Although this study examined the opinions of translation students towards HT and MTPE, some results showed that the short period of time allocated for the teaching intervention might have affected some students' opinions (namely, students' opinions about MTPE productivity). It is therefore recommended that the study is repeated over a longer period of time and that technical, temporal and cognitive efforts are examined as well (see section 6.3 for more explanations about temporal, technical and cognitive efforts).

Also, the lack of application of PE error correction apparent in MTPE TTs suggests that students may not have practiced PE skills hands-on long enough to the extent of predicting errors in the post-test. Therefore, it is recommended that a longer study with special focus on error correction is carried out for comparison of results especially in relation to errors in accuracy and grammar.

### 6.6.2 Participants and Evaluators

The results of this study were derived from comparing the PE work by female students who are native speakers of Arabic. The results could be different in case the study had been extended to males who may show different patterns of participation and interaction, or on non-native speakers of Arabic. As for sample size, having a larger sample that encompasses students from different universities and includes a gender balance may broaden the research scope and may ensure that the analysis and discussion are more convincing, while more insights into socio-cultural factors related to gender are gained. In fact, Al-Ghamdi et al. (2016) found that the learners' gender can affect their participation and interaction within an online learning space. Also, it would be interesting to undertake a similar study on three groups, i.e., a single sex female group, a single sex male group and perhaps a mixed group to explore and compare any concluded

differences. Furthermore, conducting a similar study on a larger sample size would show more significant results. This would, however, require a substantial amount of time and effort dedicated to such bigger scale research, and it is beyond the limit of this study. In addition, the translations in the present study were reviewed by two evaluators who are both native speakers of Arabic. Recruiting more evaluators would add to the reliability of the results.

### 6.6.3 MT Systems

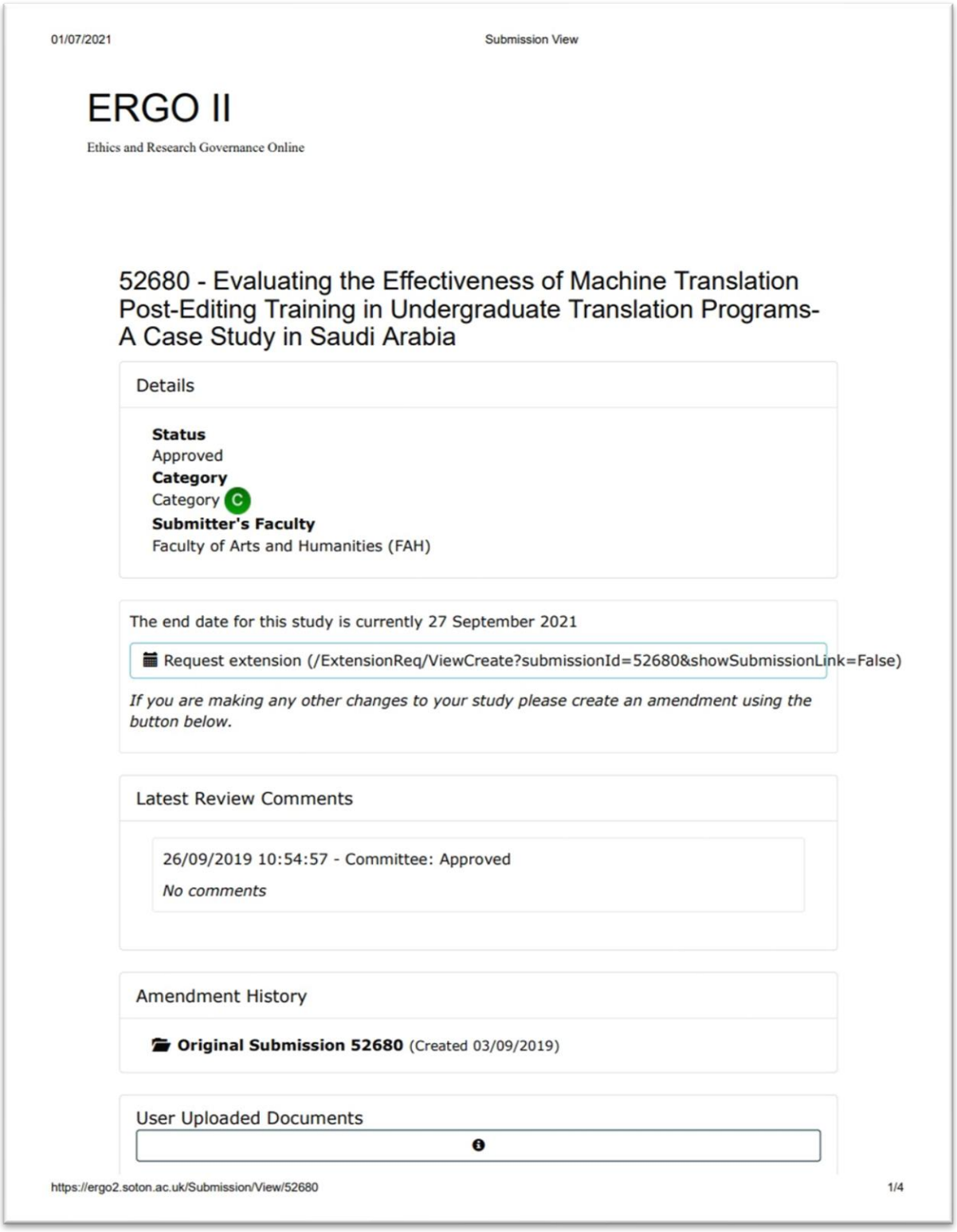
This study used Google NMT in 2019. Similar studies on updated versions of Google NMT and or other MT systems may produce different results in terms of productivity and quality (such as in the case of the more advanced and recent Amazon Translate which produces much better Arabic MT output).

### 6.6.4 Methodology

Results of the present study were derived from a mixed-method approach that aimed at evaluating a PE teaching intervention through exploring students' opinions about it, and comparing their productivity and the quality of their produced TTs with HT. It would be interesting to re-run the evaluation using more advanced productivity-measuring tool and different TQA approaches, longer texts, different text types involving both MTPE and TMs. In addition, it would also be interesting to see the effect of training students to correct the highlighted errors on their performance as suggested in the current study. Such study could be beneficial for PE training if it proved to be effective.

Combining the FGDs with observations is recommended so that more rich interpretations can be extracted from participant responses. This is especially useful in cases where participants refrain from discussing certain issues.

# Appendix A Ethical Approval- Southampton University



01/07/2021 Submission View

ERGO II

Ethics and Research Governance Online

52680 - Evaluating the Effectiveness of Machine Translation Post-Editing Training in Undergraduate Translation Programs- A Case Study in Saudi Arabia

Details

**Status**
Approved
**Category**
Category C
**Submitter's Faculty**
Faculty of Arts and Humanities (FAH)

The end date for this study is currently 27 September 2021

Request extension (/ExtensionReq/ViewCreate?submissionId=52680&showSubmissionLink=False)

*If you are making any other changes to your study please create an amendment using the button below.*

Latest Review Comments

26/09/2019 10:54:57 - Committee: Approved

*No comments*

Amendment History

**Original Submission 52680** (Created 03/09/2019)

User Uploaded Documents

https://ergo2.soton.ac.uk/Submission/View/52680 1/4

# Appendix B Focus Group Discussion Questions

### Focus Group Questions

Each participant will have a name card in front of her.

a. Readiness for MT training- How satisfied are you with the current translator training programme and its ability to train you to the market needs? Elaborate please
b. Training facilities- To what extent are you satisfied with the infrastructure (libraries, teaching rooms, language labs, access to the internet, etc.) in your college?
c. Training facilities- In general, to what extent are you satisfied with the teaching methods applied to the translation modules?
d. Readiness for MT training - Assume that the weekly schedule is 20 hours, how many hours per week of practical training on translation (Machine Translation post-editing) do you think should be offered in a bachelor's degree programme in translation?
e. MTPE- What do you know about post-editing of machine translation output?
f. RQ1.1.- Which do you think is generally more rewarding and why: human translation or machine translation post-editing?
g. RQ1.2. - How useful do you find machine translation output?
h. RQ1.3.- Which do you think is faster and why: human translation or machine translation post-editing?
i. RQ1.4.- Which do you think brings about overall better quality: human translation or machine translation post-editing? Please elaborate, why do you think so?

### Fulford (2002) Focus Group Topics

1. Profile of participants: level of study, experience with MT, training, IT familiarity, internet facilities
2. MT Uptake: MT frequency of use, PE, satisfaction with MT and PE
3. MT Training: in school training (any?)
4. Views of MT: its capabilities, potentials and viability, any scepticism (caused by complexity or fear?)
5. Choosing MT: the interest, the need for training, cases suitable for MT, time and cost of MT
6. Do you want to learn about MT?
7. MT training 'wish list': hands-on try outs, real-world tasks, guidelines for choosing the best MT system, short-term courses (for freelance translators)

# Appendix C Consent Form- English

(this consent form will be provided to the students in Arabic and the English is for ERGO team information only.)

**Study title**:

Evaluating the Effectiveness of Machine Translation Post-Editing Training in Undergraduate Translation Programs- An Exploratory Study in Saudi Arabia

**Researcher name**: Mrs. Halah Samman

**ERGO number**: 48001

***Please initial the box(es) if you agree with the statement(s):***

| | |
|---|---|
| I have read and understood the information sheet (February 2019/version no. 1 Participant Information Sheet) and have had the opportunity to ask questions about the study. | |
| I agree to take part in this research project and agree for my data to be used for the purpose of this study. | |
| I understand my participation is voluntary and I may withdraw for any reason at any point of the study. | |
| I understand that should I withdraw from the study then the information collected about me up to this point will not be used for the study. | |
| I understand that I may be quoted directly in reports of the research but that I will not be directly identified (e.g. that my name or identity will not be used). | |
| I understand that my anonymity cannot be guaranteed in these focus group discussions but that any information collected by the researchers will be kept confidential and participants will be asked to keep the discussions confidential. | |
| I understand that taking part in the study involves **audio recording and written notes** which will be transcribed for the purposes set out in the participation information sheet. | |
| I understand that my personal information collected about me such as my name will not be shared beyond the study team. | |

Name of participant (print name)……

Appendix C

Signature of participant…….

Date……

---

Name of researcher (print name) Mrs. Halah Samman ( H.Samman@soton.ac.uk )

Signature of researcher 

Date **26/2/2019**

# Appendix D Consent Form-Arabic

# نموذج موافقة

**عنوان الدراسة** :

تقييم فعالية التدريب على تحرير نصوص الترجمة الآلية في برامج الترجمة الجامعية - دراسة حالة في المملكة العربية السعودية

**اسم الباحث:** هالة سمان

**رقم ERGO** : 52680

**يرجى التوقيع في كل خانة أمام العبارة التي توافقين عليها:**

| | |
|---|---|
| لقد قرأت وفهمت محتوى ورقة المعلومات للمشارك (سبتمبر 2019 / النسخة رقم: 1 ورقة معلومات للمشارك) وقد أتيحت لي الفرصة لطرح الأسئلة عن محتوى الدراسة. | |
| أوافق على المشاركة في هذا المشروع البحثي وأوافق على استخدام بياناتي الخاصة لغرض هذه الدراسة. | |
| أتفهم أن مشاركتي تطوعية ويمكنني الانسحاب لأي سبب وفي أي مرحلة من مراحل الدراسة. | |
| أفهم أنه في حالة انسحابي من الدراسة، فإن المعلومات التي تم جمعها عني حتى هذه المرحلة لن يتم استخدامها لأغراض تحقيق أهداف الدراسة. | |
| أفهم أنه قد يتم اقتباسي مباشرة في تقارير البحث ولكن لن يتم تحديد هويتي بشكل مباشر (على سبيل المثال، لن يتم استخدام اسمي). | |
| أتفهم أنه لا يمكن ضمان سرية هويتي أثناء هذه الدراسة، ولكن أي معلومات جمعها الباحثون ستبقى سرية. | |
| أتفهم أن المشاركة في الدراسة تتضمن ملاحظات مكتوبة إلكترونية /ورقية سيتم جمعها للأغراض المحددة في ورقة معلومات المشاركة. | |
| أدرك أن معلوماتي الشخصية التي تم جمعها عني مثل اسمي لن يتم مشاركتها خارج فريق الدراسة. | |
| أوافق على تعييني في المجموعة بشكل عشوائي. | |

# Appendix E-Participant Information Sheet (PIS)-English

## Participant Information Sheet

## Focus Group Discussion

**Study Title**: Evaluating the Effectiveness of Machine Translation Post-Editing Training in Undergraduate Translation Programs- An Exploratory Study in Saudi Arabia

**Researcher name**: Mrs. Halah Samman

**ERGO number**: 48001

You are being invited to take part in the above research study. To help you decide whether you would like to take part or not, it is important that you understand why the research is being done and what it will involve. Please read the information below carefully and ask questions if anything is not clear or you would like more information before you decide to take part in this research. You may like to discuss it with others but it is up to you to decide whether or not to take part. If you are happy to participate you will be asked to sign a consent form.

**What is the research about?**

My name is Mrs. Halah Samman and I am a PhD student from the Humanities Department at Southampton University, UK, I worked as a translation lecturer at the College of Languages and Translation since 2004 until 2017, when travelled to the UK to pursue my PhD.
This is a PhD research study that aims at evaluating the effectiveness of machine translation post-editing training in undergraduate translation programs. It is expected that through investigating the students' attitudes towards machine translation and post-editing, and how they perform in a translation task (in which the speed and quality of translation produced through post-editing of machine translation is compared to those which are produced my means of conventional human translation), a complete evaluation of this type of translator training will be achieved. This project is funded by the Saudi cultural Bureau in London, UK.

**Why have I been asked to participate?**

You're eligible to take part in the focus group discussion because you have successfully finished a course on Computer-assisted Translation. There will be 6 students in each focus group discussion.

**What will happen to me if I take part?**

If you decide to take part in this study, you will participate in a one-hour focus group discussion in Arabic language that will be held in a meeting room in the college during the normal working day. The discussion group will include a total of 6 participants. It is necessary that I audio record your participation in the

discussion and then we will use the information to feed the design of a post-editing training course that will be later offered to students in the same college. However, your recorded participation will be unidentifiable in the audio recordings as your identity will only be known by the researcher, and they will be saved in a password-guarded folder on my Southampton University OneDrive-storage system.

**Are there any benefits in my taking part?**

There will be no direct benefit for you, but it is hopes that through the results of this study, future translation students will benefit from the improvement of their translator training program.

**Are there any risks involved?**

Your name will be exposed during the discussion group (your name will be written on a card placed on the table in front of you for the discussion moderator(s) and other participants to address you during the session). However, your identity will not be exposed in the written analysis or anywhere in the written PhD thesis. If you are to be mentioned, a nickname will be assigned to you.

**What data will be collected?**

Your participation in the focus group discussion will be audio recorded by me and Dr. Hind Alotaibi (If she co-moderates the focus group). Additional information about you (your level of study, and how familiar you are with machine translation, and your personal experience with machine translation) will be recorded and used for the purpose of analysing the data.

**Will my participation be confidential?**

Yes. Your participation and the information we collect about you during the course of the research will be kept strictly confidential.

Only members of the research team (Supervisors: Dr. Ian McCall, Dr. James Minney, and the researcher: Halah Samman) and responsible members of the University of Southampton may be given access to data about you for monitoring purposes and/or to carry out an audit of the study to ensure that the research is complying with applicable regulations. Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to your data. All of these people have a duty to keep your information, as a research participant, strictly confidential.

All recordings containing your voice will be unidentifiable, and they will be saved in a password-guarded folder on my Southampton University OneDrive-storage system.

**Do I have to take part?**

No, it is entirely up to you to decide whether or not to take part. If you decide you want to take part, you will need to sign a consent form to show you have agreed to take part.

**What happens if I change my mind?**

You have the right to change your mind and withdraw at any time without giving a reason and without your participant rights being affected.

If you decide to opt-out *before* that focus group discussion, please email me h.samman@soton.ac.uk stating that you no longer wish to participate in the focus group discussion.

If you withdraw from the study *after* the focus group discussion, we will not use the information obtained from you during the discussion.

**What will happen to the results of the research?**

Your personal details will remain strictly confidential. Research findings made available in any reports or publications will not include information that can directly identify you without your specific consent.

If you would like a copy of the dissertation, please email me at h.samman@soton.ac.uk

**Where can I get more information?**

If you have any questions about the study, please email me at h.samman@soton.ac.uk

**What happens if there is a problem?**

If you have a concern about any aspect of this study, you should contact the researcher who will do their best to answer your questions.

Researcher: Mrs. Halah Samman H.Samman@soton.ac.uk

If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (023 8059 5058, rgoinfo@soton.ac.uk).

**Data Protection Privacy Notice**

The University of Southampton conducts research to the highest standards of research integrity. As a publicly-funded organisation, the University has to ensure that it is in the public interest when we use personally-identifiable information about people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use information about you in the ways needed, and for the purposes specified, to conduct and complete the research project. Under data protection law, 'Personal data' means any information that relates to and is capable of identifying a living individual. The University's data protection policy governing the use of personal data by the University can be found on its website (https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page).

This Participant Information Sheet tells you what data will be collected for this project and whether this includes any personal data. Please ask the research team if you have any questions or are unclear what data is being collected about you.

Appendix E

Our privacy notice for research participants provides more information on how the University of Southampton collects and uses your personal data when you take part in one of our research projects and can be found at http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Integrity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

Any personal data we collect in this study will be used only for the purposes of carrying out our research and will be handled according to the University's policies in line with data protection law. If any personal data is used from which you can be identified directly, it will not be disclosed to anyone else without your consent unless the University of Southampton is required by law to disclose it.

Data protection law requires us to have a valid legal reason ('lawful basis') to process and use your Personal data. The lawful basis for processing personal information in this research study is for the performance of a task carried out in the public interest. Personal data collected for research will not be used for any other purpose.

For the purposes of data protection law, the University of Southampton is the 'Data Controller' for this study, which means that we are responsible for looking after your information and using it properly. The University of Southampton will keep identifiable information about you for 5 years after the study has finished after which time any link between you and your information will be removed.

To safeguard your rights, we will use the minimum personal data necessary to achieve our research study objectives. Your data protection rights – such as to access, change, or transfer such information - may be limited, however, in order for the research output to be reliable and accurate. The University will not do anything with your personal data that you would not reasonably expect.

If you have any questions about how your personal data is used, or wish to exercise any of your rights, please consult the University's data protection webpage (https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page) where you can make a request using our online form. If you need further assistance, please contact the University's Data Protection Officer (data.protection@soton.ac.uk).

**Thank you for taking the time to read the information sheet and considering taking part in the research.**

# Appendix F- Participant Information Sheet (PIS)-Arabic

**ورقة معلومات المشارك**

تقييم التدريب على تحرير الترجمة الآلية

**عنوان الدراسة**

تقييم فعالية التدريب على تحرير الترجمة الآلية في برامج الترجمة الجامعية - دراسة حالة في المملكة العربية السعودية

**اسم الباحثة**

هالة سمان

رقم ERGO 52680

عزيزتي الخريجة:

تمت دعوتك **للمشاركة في واحدة فقط** من مجموعتين في الدراسة البحثية المذكورة أعلاه (إما مجموعة التجربة أو مجموعة التحكم). لمساعدتك في تحديد ما إذا كنت ترغبين في المشاركة أم لا ، من المهم توضيح سبب إجراء البحث وما الذي سيتضمنه. يرجى قراءة المعلومات أدناه بعناية وطرح الأسئلة إذا كان أي شيء غير واضح أو كنت ترغب في مزيد من المعلومات قبل أن تقرر المشاركة في هذا البحث. قد ترغب في مناقشته مع الآخرين ، ولكن الأمر متروك لك لتقرر ما إذا كنت ستشارك أم لا. إذا كنت سعيدًا بالمشاركة ، فسيُطلب منك التوقيع على نموذج موافقة.

**ما هو موضوع البحث؟**

اسمي السيدة هالة سمان وأنا طالبة دكتوراه من قسم العلوم الإنسانية بجامعة ساوثهامبتون بالمملكة المتحدة ، وعملت محاضرة للترجمة في كلية اللغات والترجمة منذ عام 2004 حتى عام 2017 الذي سافرت فيه إلى المملكة المتحدة لمتابعة دراستي للحصول على درجة الدكتوراه .

هذه دراسة بحثية لدرجة الدكتوراه تهدف إلى تقييم فعالية تحرير الترجمة الآلية في برامج الترجمة الجامعية. من المتوقع من خلال دراسة آراء الطلاب بتحرير الترجمة الآلية وكيفية أدائهم في مهمة ترجمة (تتم مقارنة سرعة وجودة الترجمة التي يتم إنتاجها من خلال تحرير الترجمة الآلية بتلك التي يتم إنتاجها عبر الترجمة البشرية التقليدية) ، سيتم إجراء تقييم كامل لهذا النوع من التدريب على المترجمين.

يتم تمويل هذا المشروع من قبل المكتب الثقافي السعودي في لندن ، المملكة المتحدة.

**لماذا طلبت مني المشاركة؟**

أنت مؤهلة للمشاركة في هذا التقييم للتدريب على تحرير الترجمة الآلية لأنك قد أنهيت دورة تدريبية عن الترجمة بمساعدة الكمبيوتر بنجاح. سيكون هناك ما مجموعه 60 طالبًا في التقييم الشامل.

**ماذا سيحدث لي إذا شاركت؟**

بمجرد الموافقة على المشاركة في الدراسة ، سيتم تخصيص مكان عشوائي لك في مجموعة (ستحقق كلتا المجموعتين نفس أهداف المادة ولكن أساليب التدريس مختلفة). بمجرد بدء الدراسة ، من المتوقع منك:

(1) حضور دورة تدريبية مدتها 5 أسابيع (ساعات الاتصال: ما مجموعه 10 ساعات (ساعتان في الأسبوع)) في الترجمة العلمية والتقنية (تتضمن مهام أسبوعية لتعزيز مهارة الترجمة)

سيتم منحك واجبات أسبوعية في المختبر لتعزيز المهارة (4 نصوص: نص واحد / أسبوع - الأسبوع الأول بدون واجب). سيُطلب منك تقديم النسخة النهائية للترجمة بنهاية الجلسة. سوف تستخدمين تقنية الترجمة التي تعلمتها في الفصل.

(2) حضور اختبارين في الترجمة. واحد في بداية الدورة وواحد بعد الانتهاء من الدورة. ستستغرق كل جلسة ما يصل إلى ساعتين ، ولكن سيُطلب منك تقديم الترجمة النهائية في أسرع وقت ممكن. ستخدم هذه المهام أغراض البحث فقط ولن يتم احتسابها في الدرجات الرسمية.

(3) يمكن تسجيل ملاحظة تفاعلك / حافزك في الفصل الدراسي في دفتر يوميات المعلم.

(4) ملء استطلاع واحد (20 دقيقة) حول الدورة التدريبية.

إذا انضمت إلى **أي مجموعة** ، فستكون مشاركتك مجهولة عبر اتخاذ هذه التدابير الثلاثة:

• لن يتم تحديد جميع المعلومات الشخصية المحددة في المهام / التجربة / الاستبيان. ستتم إزالة جميع أسماء الطلاب ومعرفاتهم وأية بيانات شخصية أخرى واستبدالها بأسماء مستعارة ورموز.

• سيتم تخزين سرعة الترجمة المحسوبة (من خلال استخدام برنامج يسمى Time Doctor) - على (سيرفر Southampton University OneDrive).

• سيتم تخزين وثائق التجربة (النصوص المترجمة) على (سيرفر Southampton University OneDrive). لن يتم مشاركة البيانات مع طرف ثالث إلا إذا طلبت جامعة ساوثهامبتون ذلك.

• سيتم حفظ الاستبانات (من مجموعة التجربة فقط) ونماذج الموافقة الموقعة (من المجموعتين) بشكل آمن في خزانة ملفات مغلقة.

**هل هناك أي فائدة من مشاركتي؟**

نعم. عندما تكملين (1) حضور الدورة ، (2) تشاركين في التجربة ، و (3) تملئين الاستبيان ، ستتلقي:

- شهادة رسمية من الجامعة بحضور دورة تحرير الترجمة الآلية.

ومن المأمول أيضًا أن يستفيد طلاب الترجمة في المستقبل من خلال مشاركتك ونتائج هذه الدراسة من أجل تحسين برنامج تدريس الترجمة.

**هل هناك أي مخاطر تنطوي عليها المشاركة؟**

سيتم الكشف عن اسمك وهويتك أثناء التجربة (أيضًا في الدورة التدريبية إذا كنت في مجموعة التجربة) ، وقد يقوم مشاركون آخرون بمخاطبتك خلال هذه الجلسات. ومع ذلك ، لن يتم كشف هويتك في التحليل المكتوب أو في أي مكان في أطروحة الدكتوراه المكتوبة. إذا تم ذكر اسمك ، فسيتم تخصيص اسم وهمي للدلالة عليك.

**ما هي البيانات التي سيتم جمعها؟**

**خلال الدورة:**

سيتم تسجيل ملاحظة تفاعلك في الفصل في دفتر يوميات المعلم (على سبيل المثال ، سيقوم المعلم بكتابة ملاحظاتها حول تفاعلك في الفصل).

**في التجربة:**

• بيانات شخصية

اسم الطالب ، والعمر ، ومستوى الخبرة الدراسية مع MT ، والخبرة التدريبية ، وألفة تكنولوجيا المعلومات.

• بيانات البحث (التجربة)

سرعة الطلاب المحسوبة

+ جودة النصوص المترجمة

**في الاستبيان:**

• البيانات الشخصية

ملف تعريف المشاركين: الاسم والعمر ومستوى الدراسة ، والخبرة في مجال التدريب ، والخبرة التدريبية ، وإلمام تكنولوجيا المعلومات ، والآراء حول مرافق الإنترنت في الكلية.

• بيانات البحوث

مسح ورقي حول تجارب الطلاب وموقفهم من الترجمة الآلية ، والتحرير اللاحق ، والترجمة البشرية + ملاحظات حول تفاعلك / حافزك في الفصل (فقط إذا لزم الأمر).

**هل ستكون مشاركتي سرية؟**

نعم. سيتم الاحتفاظ بمشاركتك والمعلومات التي نجمعها عنك أثناء البحث في سرية تامة.

يمكن فقط لأعضاء فريق البحث (المشرفون: الدكتور إيان مكول ، والدكتور جيمس ميني ، والباحث: هالة سمان) والأعضاء المسؤولون في جامعة ساوثهامبتون الوصول إلى البيانات الخاصة بك لأغراض المراقبة و / أو القيام بها التدقيق في الدراسة للتأكد من أن البحث يتوافق مع اللوائح المعمول

بها. قد يحتاج الأفراد من السلطات التنظيمية (الأشخاص الذين يتحققون من إجراء الدراسة بشكل صحيح) إلى الوصول إلى بياناتك. كل هؤلاء الأشخاص عليهم واجب الحفاظ على سرية معلوماتك ، بصفتك مشاركًا في الأبحاث.
جميع أوراق الترجمة التي تحتوي على مشاركتك ستكون غير قابلة للتعريف ، وسيتم حفظها في مجلد محمي بكلمة مرور على نظام التخزين بجامعة ساوثهامبتون في OneDrive.

**هل يجب علي المشاركة؟**
لا ، الأمر متروك لك تمامًا لتقرر ما إذا كنت ستشارك أم لا. إذا قررت رغبتك في المشاركة ، فستحتاج إلى توقيع نموذج موافقة لإظهار موافقتك على المشاركة.

**ماذا يحدث إذا غيرت رأيي؟**
لديك الحق في تغيير رأيك والانسحاب في أي وقت دون إبداء سبب ودون أن تتأثر حقوق المشارك الخاصة بك.
إذا قررت إلغاء الاشتراك قبل بدء التقييم ، فيرجى إرسال رسالة بريد إلكتروني إلي على h.samman@soton.ac.uk تفيد أنك لم تعد ترغب في المشاركة.
إذا انسحبت من الدراسة بعد الدورة التدريبية أو التجربة أو الاستطلاع ، فلن نستخدم المعلومات التي حصلت عليها خلال هذا التقييم.

**ماذا سيحدث لنتائج البحث؟**
ستبقى بياناتك الشخصية سرية للغاية. لن تتضمن نتائج البحوث المتاحة في أي تقارير أو منشورات أية معلومات يمكن أن تحدد هويتك مباشرة دون موافقتك المحددة.
إذا كنت ترغب في الحصول على نسخة من الرسالة ، يرجى مراسلتي عبر البريد الإلكتروني على h.samman@soton.ac.uk

**أين يمكنني الحصول على مزيد من المعلومات؟**
إذا كان لديك أي أسئلة حول الدراسة ، يرجى مراسلتي عبر البريد الإلكتروني على h.samman@soton.ac.uk

**ماذا يحدث إذا كان هناك مشكلة؟**
إذا كنت مهتمًا بأي جانب من جوانب هذه الدراسة ، فيجب عليك الاتصال بالباحث الذي سيبذل قصارى جهده للإجابة على أسئلتك.
الباحثة: السيدة هالة سمان H.Samman@soton.ac.uk
إذا كنت لا تزال غير سعيد أو لديك شكوى بشأن أي جانب من جوانب هذه الدراسة ، فيرجى الاتصال بمدير النزاهة والحكم في جامعة ساوثهامبتون (023 8059 5058 ، rgoinfo@soton.ac.uk).

**إشعار خصوصية حماية البيانات**
تجري جامعة ساوثهامبتون أبحاثًا وفقًا لأعلى معايير النزاهة البحثية. بصفتها منظمة ممولة من القطاع العام ، يتعين على الجامعة التأكد من أنها تخدم المصلحة العامة عندما نستخدم معلومات التعريف الشخصية عن الأشخاص الذين وافقوا على المشاركة في البحث. هذا يعني أنه عندما توافق على المشاركة في دراسة بحثية ، سنستخدم معلومات عنك بالطرق اللازمة ، وللأغراض المحددة ، لإجراء وإكمال مشروع البحث. بموجب قانون حماية البيانات ، تعني "البيانات الشخصية" أي معلومات تتعلق بقادر حي على الفرد وقادر على تحديدها. يمكن الاطلاع على سياسة الجامعة لحماية البيانات التي تحكم استخدام الجامعة للبيانات الشخصية على موقعها على الإنترنت
( https://www.southampton.ac.uk/legalservices/what-we-do/data-protection-and-foi.page).
تخبرك ورقة معلومات المشاركين هذه بالبيانات التي سيتم جمعها لهذا المشروع وما إذا كان هذا يتضمن أي بيانات شخصية. يرجى سؤال فريق البحث إذا كان لديك أي أسئلة أو غير واضح ما هي البيانات التي يتم جمعها عنك.
يوفر إشعار الخصوصية للمشاركين في الأبحاث مزيدًا من المعلومات حول كيفية قيام جامعة ساوثامبتون بجمع واستخدام بياناتك الشخصية عندما تشارك في أحد مشاريعنا البحثية ويمكن العثور عليها على الموقع http://www.southampton.ac.uk/assets/ شير / الإنترانت / لير ة سورية / العامة / بحوث٪ and20٪ Integrity20٪ Privacy20٪ Notice20 / الخصوصية٪ Notice20٪ for20٪ Research20٪ Participants.pdf20

سيتم استخدام أي بيانات شخصية نجمعها في هذه الدراسة فقط لأغراض إجراء أبحاثنا وسيتم التعامل معها وفقًا لسياسات الجامعة وفقًا لقانون حماية البيانات. إذا تم استخدام أي بيانات شخصية يمكن التعرف عليها مباشرة ، فلن يتم الكشف عنها لأي شخص آخر دون موافقتك إلا إذا كانت جامعة ساوثهامبتون مطالبة قانونًا بالكشف عنها.

يتطلب قانون حماية البيانات أن يكون لدينا سبب قانوني صحيح ("الأساس القانوني") لمعالجة واستخدام بياناتك الشخصية. الأساس القانوني لمعالجة المعلومات الشخصية في هذه الدراسة البحثية هو لأداء مهمة تم تنفيذها لتحقيق المصلحة العامة. لن يتم استخدام البيانات الشخصية التي يتم جمعها للبحث لأي غرض آخر.

لأغراض قانون حماية البيانات ، فإن جامعة ساوثهامبتون هي "مراقب البيانات" لهذه الدراسة ، مما يعني أننا مسؤولون عن رعاية معلوماتك واستخدامها بشكل صحيح. ستحتفظ جامعة ساوثهامبتون بمعلومات تعريفية عنك لمدة 5 سنوات بعد انتهاء الدراسة وبعد ذلك سيتم إزالة أي رابط بينك وبين معلوماتك.

لحماية حقوقك ، سنستخدم الحد الأدنى من البيانات الشخصية اللازمة لتحقيق أهداف الدراسة البحثية. ومع ذلك ، قد تكون حقوق حماية البيانات الخاصة بك - مثل الوصول إلى هذه المعلومات أو تغييرها أو نقلها - محدودة ، حتى يكون مخرجات البحث موثوقة ودقيقة. لن تفعل الجامعة أي شيء مع بياناتك الشخصية التي لا تتوقعها بشكل معقول.

إذا كانت لديك أي أسئلة حول كيفية استخدام بياناتك الشخصية ، أو ترغب في ممارسة أي من حقوقك ، فيرجى الرجوع إلى صفحة الويب الخاصة بحماية البيانات في الجامعة (https://www.southampton.ac.uk/legalservices/what-we-do/ data-protection-and-foi.page) حيث يمكنك تقديم طلب باستخدام نموذجنا على الإنترنت. إذا كنت بحاجة إلى مزيد من المساعدة ، فيرجى الاتصال بمسؤول حماية البيانات بالجامعة (data.protection@soton.ac.uk).

**شكرا جزيلا لاستثمارك الوقت الكافي لقراءة ورقة المعلومات والتفكير في المشاركة في البحث.**

# Appendix G -External Approval

جامعة الملك سعود (034)
هاتف 87 18 805 11 966+
فاكس 56 45 805 11 966+

المملكة العربية السعودية
ص.ب 7695 الرياض 11472
www.ksu.edu.sa

جامعة الملك سعود
King Saud University

أقسام العلوم الإنسانية

Dear Ms. Halah M. Samman,

I am pleased to tell you that you have been granted approval to collect data for your Ph.D. research, entitled "The effectiveness of machine translation post-editing: An empirical study on translation students in Saudi Arabia", at the College of Languages and Translation, King Saud University, Riyadh. This approval allows you to use the College classrooms, labs, lab facilities, software, and to approach faculty members and instructors for classroom attendance and giving lessons. We wish you a successful data collection trip and we are looking forward to see you here in Riyadh.

All the best,

Reem Salem

Vice Dean for Female Students' Affairs,

College of Languages and Translation,

King Saud University, Riyadh.

ralsalem@ksu.edu.sa

# Appendix H - Visited Research Designs

**MAIN Question:** How effective is MTPE training in an undergraduate translation programme in Saudi Arabia?

| **Plan 1** | **Plan 2** | **Plan 3** |
|---|---|---|
| Worldview: **Pragmatic: Constructivist -->Postpositivist** (pre-post intervention)<br>Approach: **mixed** (**Sequential Exploratory Design**)<br>Methods:<br>1. (April,2019) **focus group** to seek students' opinions and needs + intervention<br>***RQ:*** *How do students feel about MTPE?*<br>***RQ:*** *What are the needs of students regarding MTPE training?*<br>2. (Spet,2019) Experiment<br>(a) keystroke logging (task speed)<br>(b) TQA by reviewers<br>***RQ:*** *What is the effect of MTPE training on students' task speed?*<br>***RQ:*** *What is the effect of MTPE training on students' TQ?*<br>***RQ:*** *What is the correlation between MTPE training and (Speed/Quality)?*<br>***Possible RQ:*** *What is the correlation between student's attitudes and their performance (speed/TQ) in the task?* | Worldview: **Postpositivist** (post intervention)<br>Approach: Quantitative<br>Methods:<br>1. (Sept 2019) Closed questionnaire to seek opinions + levels of students + intervention<br>2. (Oct 2019) experiment (randomised control design)<br>(a) keystroke logging (task speed)<br>(b) TQA by reviewers<br>***RQ:*** *What is the effect of MTPE training on students' task speed?*<br>***RQ:*** *What is the effect of MTPE training on students' TQ?*<br>***RQ:*** *What is the correlation between MTPE training and (Speed/Quality)?*<br>***Possible RQ:*** *What is the correlation between student's attitudes/levels and their performance (speed/TQ) in the task?* | Worldview: **Constructivist**<br>Approach: Qualitative<br>Methods:<br>1. (April, 2019) **focus group** to seek opinions and training needs + intervention<br>***RQ:*** *How do students feel about MTPE?*<br>***RQ:*** *What are the needs of students regarding MTPE training?*<br>2. (Apr-May, 2019) (effects) **Class observation**<br>3. (Apr-May, 2019) retrospective **interviews** (Patton's "maximum variation sampling")<br>***RQ:*** *What are the perceived effects of MTPE on students*<br>***Possible RQ:*** *What is the correlation between student's attitudes and their perceived effects of MTPE training?* |

| **Strengths:** | **Strengths:** | **Strengths:** |
|---|---|---|
| -Closest approach to answering my research question (through its sub-research questions). | -One treatment is directly compared to another to establish superiority. | -All the problems and the topics covered under this research are in detail. |

| | | |
|---|---|---|
| -Quan component can make qual approach more acceptable to quan-biased audience.<br>- It has broader focus than the single-method design, thus it gathers more data in different modes about the phenomenon.<br>- It can provide insight into the complexity of a social phenomenon by producing findings that illustrate that complexity.<br>-The breadth of its findings that brings value to the research process itself by highlighting the shortcomings in each of the methods that were used, and ways used to compensate for them. | -This design can make causal inferences, i.e., it is the **strongest empirical evidence of a treatment's efficacy**<br>-**Minimises bias**<br>Randomisation minimises allocation bias and selection bias<br>-Blinding minimises performance bias<br>Double blinding minimises assessment bias<br>-Allocation concealment minimises both performance and assessment bias<br>-Prospective design minimises recall error and selection bias<br>-**Minimises confounding factors**<br>-Randomisation minimises confounding due to unequal distribution of prognostic factors<br>Randomisation makes groups comparable according both known and unknown factors<br>-Blocked randomisation makes groups comparable within known confounding factors<br>-**Statistical reliability**<br>Statistical test of significance is readily interpretable when the study is randomised<br>-Sample size (when adequately powered) avoids both Type 1 error (where the null hypothesis is incorrectly rejected) and Type 2 error (where the null hypothesis is incorrectly accepted) | -This method majorly focuses on small groups which ultimately do not require more expenses when compared to quantitative research.<br>-On the emergence of new developed information and findings, the revision, direction and framework of the data can be done easily quickly.<br>-The data are collected from a small group which bounds it to be universal for a large population.<br>-The data with this method is collected based on genuine efforts and gives a clear vision on what can be expected. |

# Appendix I – Experiment Texts (Ta and Tb)

## Ta (pre-test):

If your showerhead is clogged with years of mineral deposit build-up, it may need a thorough cleansing. Read this article to learn two simple ways of cleaning your showerhead using vinegar and water.

1. **Gather your supplies.** One way to clean your showerhead is by taking it off the pipe and soaking it in vinegar. Here is what you will need for this method:

    - Pot, bucket, or another container large enough to fit the showerhead
    - Distilled white vinegar, Wrench and old rag (optional)
    - An old toothbrush, Soft cloth, such as microfiber.

2. **Remove the showerhead by twisting it counter-clockwise.** If you are having difficulties twisting it, try wrapping an old rag around the connecting nut, and then twisting it with a wrench. The rag will help protect the surface of your showerhead.
3. **Put the showerhead into a pot.** Consider using a smaller container that just about fits the showerhead; this way, you will use less vinegar. You can also use a small bucket or plastic bin.
4. **Fill the pot with enough white vinegar to cover the showerhead.** The acids in the vinegar will help dissolve the white mineral deposits on the showerhead.
5. **Let the showerhead soak in the vinegar for 30 minutes to overnight** as the more soiled the showerhead is, the longer you will have to leave it in the vinegar.
6. **Take the showerhead out of the pot and rinse it.** You should see the mineral deposits flaking off.

---

## Tb (post-test):

The first rule of using any washing machine is using the right laundry detergent. Ariel Orignal Liquid laundry detergent or Ariel Original Powder Laundry Detergent are both great choices for most loads, giving you great stain removal in 1 wash. So just follow the tips below to get a perfect wash with your front-loading washing machine.

**Using a front load washer:**

**Dose correctly**

Follow the instructions on your detergent's box or bottle. You will need more detergent for dirtier clothes, more laundry or hard water.

**Load your laundry**

Place your dirty laundry into the drum of the washing machine. Make sure to leave enough space for the clothes to move around in the wash. Do the 1-palm trick: if your palm fits between the clothes and the machine, it means that you have loaded the machine correctly. If your palm doesn't fit, take some garments out.

**Add fabric conditioner to the dispenser drawer**

Make sure not to fill past the MAX line, otherwise it may not dispense at all.

**Choose the temperature**

Select the right temperature based upon the fabric care labels on your clothes. If there are multiple temperatures, choose the lowest one: Ariel performs just as well in a cold wash. For more info on fabric care labels, click here.

**Select the right spin cycle**

Again, check the fabric care labels on your clothes. As a rule, delicates need a slower cycle, while cotton and other durable fabrics will be better with a faster cycle.

**List of References**

Abdelaal, N.M. and Alazzawie, A., 2020. Machine Translation: The Case of Arabic-English Translation of News Texts. *Theory & Practice in Language Studies*, *10*(4).

Abu-Ayyash, E.A., 2017. Errors and non-errors in English-Arabic machine translation of gender-bound constructs in technical texts. *Procedia Computer Science*, *117*, pp.73-80.

Abu-ghararah, B., 2017. The Gap between Translator Training and the Translation Industry in Saudi Arabia. *AWEJ for translation & Literary Studies Volume*, *1*.

Ajzen, I., 2005. *Attitudes, personality and behaviour*. McGraw-Hill Education (UK).

Alanazi, M.S., 2019. *The Use of Computer-Assisted Translation Tools for Arabic Translation: User Evaluation, Issues, and Improvements*. Kent State University.

Albarracin, D. and Shavitt, S., 2018. Attitudes and attitude change. *Annual review of psychology*, *69*, pp.299-327.

Al-Batineh, M. and Bilali, L., 2017. Translator training in the Arab world: are curricula aligned with the language industry?. *The Interpreter and Translator Trainer*, *11*(2-3), pp.187-203.

Alenezi, A., 2016. *Development of Translation Curricula at Undergraduate Translation Courses in Saudi Universities: Exploring Student Needs and Market Demands* (Doctoral dissertation, University of Leicester).

Al-Gahtani, S.S., Hubona, G.S. and Wang, J., 2007. Information technology (IT) in Saudi Arabia: Culture and the acceptance and use of IT. *Information & management*, *44*(8), pp.681-691.

Al Ghamdi, A., Samarji, A. and Watt, A., 2016. Essential considerations in distance education in KSA: Teacher immediacy in a virtual teaching and learning environment. *International Journal of Information and Education Technology*, *6*(1), p.17.

Alhaisoni, E. and Alhaysony, M., 2017. An investigation of Saudi EFL university students' attitudes towards the use of Google Translate. *International Journal of English Language Education*, *5*(1), pp.72-82.

Alhihi, N., 2016. Lexical problems in English to Arabic translation: A critical analysis of health documents in Australia. *Arab World English Journal (AWEJ) Vol*, *6*.

Al-Jarf, R., 1999. Unemployed female translators in Saudi Arabia: causes and solutions. *Meta: journal des traducteurs/Meta: Translators' Journal*, *44*(2), pp.391-397.

Al-Jarf, R., 2017. Technology integration in translator training in Saudi Arabia. *International Journal of Research in Engineering and Social Sciences*.

Al-khresheh, M.H. and Almaaytah, S.A., 2018. English Proverbs into Arabic through Machine Translation. *International Journal of Applied Linguistics and English Literature*, *7*(5), pp.158-166.

Allen, J., 2003. Post-editing. *Benjamins Translation Library*, *35*, pp.297-318.

Alliger, G.M. and Janak, E.A., 1989. Kirkpatrick's levels of training criteria: Thirty years later. *Personnel psychology*, *42*(2), pp.331-342.

Al Lily, A.E., 2011. On line and under veil: Technology-facilitated communication and Saudi female experience within academia. *Technology in Society*, *33*(1-2), pp.119-127.

Al Mahasees, Z., 2020. Diachronic Evaluation of Google Translate, Microsoft Translator and Sakhr in English-Arabic Translation. *Unpublished Master's Thesis, the University of Western Australia, Australia*.

Almutawa, F. and Izwaini, S., 2015. Machine translation in the Arab world: Saudi Arabia as a case study. *Trans-Kom. Wissenschaftliche Zeitschrift Für Translation Und Kommunikation*, *8*(2), pp.382-414.

Al-Otaibi, A.A., 2015. *Saudi Translation Project (2000-2015): A Historiographical Bibliometric Study* (Doctoral dissertation).

Alotaibi, H.M., 2014. Teaching CAT Tools to Translation Students: an Examination of Their Expectations and Attitudes. *Arab World English Journal*.

Alsalem, R., 2019. The Effects of the Use of Google Translate on Translation Students' Learning Outcomes. *AWEJ for Translation & Literary Studies, 3*(4), pp. 46-60.

Al-Samawi, A.M., 2014. Language errors in machine translation of encyclopedic Texts from English into Arabic: the case of Google Translate. *Arab World English Journal*, pp.182-211.

Alqudsi, A., Omar, N. and Shaker, K., 2014. Arabic machine translation: a survey. *Artificial Intelligence Review*, *42*(4), pp.549-572.

Ameur, M.S.H., Meziane, F. and Guessoum, A., 2020. Arabic machine translation: A survey of the latest trends and challenges. *Computer Science Review*, *38*, p.100305.

Anderson, V., 2007. The value of learning: a new model of value and evaluation. *Chartered Institute of Personnel and Development*.

Aranberri, N., Labaka, G., de Ilarraza, A.D. and Sarasola, K., 2014. Comparison of post-editing productivity between professional translators and lay users. In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 20-33).

Arenas, A.G., Depraetere, H. and O'Brien, S., 2012. What we know and what we would like to know about post-editing. *Revista Tradumàtica: tecnologies de la traducció*, (10), pp.211-218.

Baki, R., 2004. Gender-Segregated Education in Saudi Arabia: Its Impact on Social Norms and the Saudi Labor Market. *Education policy analysis archives*, *12*(28), p.n28.

Berrichi, S. and Mazroui, A., 2018, October. Benefits of morphosyntactic features on English-Arabic statistical machine translation. In *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)* (pp. 244-248). IEEE.

Borg, W., Gall, M.D. and Gall, M.D., 2009. *Research in education: An introduction.* Longman.

Bersin, J., 2005. The four stages of e-learning. *What Works*, pp.2-56.

Bowker, L., 2005. Productivity vs quality. *A Pilot Study on the impact of translation memory systems. Localization Reader*, pp.133-140.

Bowker, L. and Ehgoetz, M., 2007. Exploring user acceptance of machine translation output: A recipient evaluation. *Across boundaries: International perspectives on translation studies*, pp.209-224.

Breckler, S.J. and Wiggins, E.C., 1989. On defining attitude and attitude theory: Once more with feeling. *Attitude structure and function*, pp.407-427.

Bree, R.T. and Gallagher, G., 2016. Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach. *All Ireland Journal of Higher Education*, *8*(2).

Brinkerhoff, R., 2003. *The success case method: Find out quickly what's working and what's not*. Berrett-Koehler Publishers.

Brosnan, M.J., 2002. *Technophobia: The psychological impact of information technology*. Routledge.

Bundgaard, K., 2017. Translator Attitudes towards Translator-Computer Interaction–Findings from a Workplace Study. *Hermes*, (56), pp.125-144.

Cadwell, P., Castilho, S., O'Brien, S. and Mitchell, L., 2016. Human factors in machine translation and post-editing among institutional translators. *Translation Spaces*, *5*(2), pp.222-243.

Cameron, R., 2009. A sequential mixed model research design: Design, analytical and display issues. *International journal of multiple research approaches*, *3*(2), pp.140-152.

Carl, M., Dragsted, B., Elming, J., Hardt, D. and Jakobsen, A.L., 2011. The process of post-editing: a pilot study. *Copenhagen Studies in Language*, *41*, pp.131-142.

Castilho, S., Doherty, S., Gaspari, F. and Moorkens, J., 2018. Approaches to human and machine translation quality assessment. In *Translation Quality Assessment* (pp. 9-38). Springer, Cham.

Cavalli-Sforza, V. and Phillips, A.B., 2012. Using morphology to improve Example-Based Machine Translation. *Challenges for Arabic Machine Translation*, *9*, p.23.

Çetiner, C., 2018. Analyzing the attitudes of translation students towards cat (computer-aided translation) tools. *Journal of Language and Linguistic Studies*, *14*(1), pp.153-161.

Chang, R. and Little, T.D., 2018. Innovations for evaluation research: Multiform protocols, visual analog scaling, and the retrospective pre-test –post-test design. *Evaluation & the health professions*, *41*(2), pp.246-269.

Cherif, N.B.J., 2021. The impact of training on productivity and wages in Tunisia. *International Journal of Manpower*.

Cohen, J., 1992. A power primer. *Psychological bulletin*, *112*(1), p.155.

Cohen, J., 2013. *Statistical power analysis for the behavioral sciences*. Academic press.

Cohen, L., Manion, L. and Morrison, K., 2013. *Research methods in education*. Routledge.

Coloma, R.S., 2008. Border crossing subjectivities and research: Through the prism of feminists of color. *Race Ethnicity and Education*, *11*(1), pp.11-27.

Colombo, E. and Stanca, L., 2014. The impact of training on productivity: evidence from a panel of Italian firms. *International Journal of Manpower*.

Condon, S., Parvaz, D., Aberdeen, J., Doran, C., Freeman, A. and Awad, M., 2010. *Evaluation of machine translation errors in English and Iraqi Arabic*. MITRE CORP MCLEAN VA.

Creswell, J.W., 2003. A framework for design. *Research design: Qualitative, quantitative, and mixed methods approaches*, pp.9-11.

Creswell, J.W., 2014. *Qualitative, quantitative and mixed methods approaches*. Sage.

Creswell, J.W. and Clark, V.L.P., 2017. *Designing and conducting mixed methods research*. Sage.

Čulo, O. and Nitzke, J., 2016. Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation* (pp. 106-114).

Daems, J., 2016. *A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude* (Doctoral dissertation, Ghent University).

Daems, J., Carl, M., Vandepitte, S., Hartsuiker, R. and Macken, L., 2016. The effectiveness of consulting external resources during translation and post-editing of general text types. In *New Directions in Empirical Translation Process Research* (pp. 111-133). Springer, Cham.

Daems, J., Macken, L. and Vandepitte, S., 2013, September. Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+ PE. In *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice* (Vol. 2, pp. 63-71).

Daems, J., Vandepitte, S., Hartsuiker, R. and Macken, L., 2017. Translation methods and experience: A comparative analysis of human translation and post-editing with students and professional translators. *Meta: Journal des traducteurs/Meta: Translators' Journal*, *62*(2), pp.245-270.

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pp.319-340.

De Almeida, G. and O'Brien, S., 2010, May. Analysing Post-Editing Performance: Correlations with Years of Translation Experience. In *EAMT*.

De Almeida, G., 2013. *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages* (Doctoral dissertation, Dublin City University).

De Boer, M.R., Waterlander, W.E., Kuijper, L.D., Steenhuis, I.H. and Twisk, J.W., 2015. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), pp.1-8.

Depraetere, I., 2010, May. What counts as useful advice in a university post-editing training context? Report on a case study. In *Annual conference of the European Association for Machine Translation,* (pp. en-ligne). http://www. mt-archive. info/EAMT-2010-Depraetere-2. pdf.

Dillon, A., 2001. User acceptance of information technology. London: Taylor and Francis.

Doherty, S. and Kenny, D., 2014. The design and evaluation of a statistical machine translation syllabus for translation students. *The Interpreter and Translator Trainer*, *8*(2), pp.295-315.

Doherty, S. and Moorkens, J., 2013. Investigating the experience of translation technology labs: pedagogical implications. *Journal of Specialised Translation*, *19*, pp.122-136.

Downes, A. (2015). *An Overview: Brinkerhoff's Method for Learning Evaluation*. Available at: https://www.watershedlrs.com/blog/overview-brinkerhoff-learning-evaluation/ (Accessed: 8 March 2019).

Drennan, J. and Hyde, A., 2008. Controlling response shift bias: the use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education*, *33*(6), pp.699-709.

Duff, P., 2018. *Case study research in applied linguistics*. Routledge.

Elsherif, H.M. and Soomro, T.R., 2017. Perspectives of arabic machine translation. *Journal of Engineering Science and technology*, *12*(9), pp.2315-2332.

Ericsson, K.A. and Simon, H.A., 1998. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), pp.178-186.

Eser, O., 2015. A model of translator's competence from an educational perspective. *International Journal of Comparative Literature and Translation Studies*, *3*(1), pp.4-15.

Farghal, M., 1995. Lexical and discoursal problems in English-Arabic translation. *Meta: Journal Des Traducteurs / Meta: Translators' Journal*, *40* (1), pp. 54-61.

Fatani, A., 2007. Globalisation and the translation industry in Saudi Arabia: factors inhibiting the use and integration of translation technology in to mainstream coursework. In *Proceedings of the UNESCO Forum on Higher Education. 2nd Regional Seminar for Arab States* (pp. 24-25).

Fatani, A., 2009. The state of the translation industry in Saudi Arabia. *Translation Journal*, *13*(4), pp.1-8.

Fiederer, R. and O'Brien, S., 2009. Quality and machine translation: A realistic objective. *The journal of Specialised translation*, *11*(11), pp.52-74.

Fife, E.M., 2007. Using focus groups for student evaluation of teaching. *MountainRise Int J Scholarsh Teach Learn*, *4*, pp.1-19.

Flesch, R., 1948. A new readability yardstick. *Journal of applied psychology*, *32*(3), p.221.

Fulford, H., 2002. Freelance translators and machine translation: An investigation of perceptions, uptake, experience and training needs. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*.

Gabr, M., 2001. Toward a model approach to translation curriculum development. *Translation Journal*, *5*(2), pp.1-13.

Gadamer, H.G., 1960. " Hinweise auf:" Möller, Absurdes Sein Pareyson, Estetica. *Philosophische Rundschau*, *8*(2/3).

Garcia, I., 2011. Translating by post-editing: is it the way forward?. *Machine Translation*, *25*(3), p.217.

Gaspari, F., Toral, A., Naskar, S.K., Groves, D. and Way, A., 2014, October. Perception vs reality: Measuring machine translation post-editing productivity. In *Third Workshop on Post-Editing Technology and Practice* (Vol. 60).

Gaspari, F., Almaghout, H. and Doherty, S., 2015. A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, *23*(3), pp.333-358.

Gentzler, E., 2001. *Contemporary translation theories* (Vol. 21). Multilingual Matters.

Ghasemi, A. and Zahediasl, S., 2012. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), p.486.

Ghazala, H.S., 2012. *Translation as problems and solutions: A textbook for university students and trainee translators*. Konooz Al-Marefa.

Giddings, L.S. and Grant, B.M., 2006. Mixed methods research for the novice researcher. *Contemporary nurse*, *23*(1), pp.3-11.

Görög, A. (2015). *Translation Productivity Revisited*, [Blog] Available at: >https://blog.taus.net/translation-productivity-revisited< (Accessed 16 April 2021).

Gouldthorpe, J.L. and Israel, G.D., 2013. Capturing Change: Comparing Pretest-Posttest and Retrospective Evaluation Methods: WC135, 1/2013. *Edis*, *2013*(1).

Graham, M., Milanowski, A. and Miller, J., 2012. Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings. *Online Submission*.

Gravetter, F.J. and Forzano, L.A.B., 2018. *Research methods for the behavioral sciences*. Cengage Learning.

Green, S., Heer, J. and Manning, C.D., 2013, April. The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 439-448).

Guerberof, A., 2009. Productivity and quality in MT post-editing. In *Beyond Translation Memories: New Tools for Translators Workshop*.

Guerberof Arenas, A., 2013. What do professional translators think about post-editing. *JoSTrans The journal of specialised translation*, *19*, pp.75-95.

Guerberof Arenas, A., 2013. What do professional translators think about post-editing. *JoSTrans The journal of specialised translation*, *19*, pp.75-95.

Guerberof Arenas, A. and Moorkens, J., 2019. Machine translation and post-editing training as part of a master's programme. *Jostrans: The Journal of Specialised Translation*, (31), pp.217-238.

Guest, G., Namey, E. and McKenna, K., 2017. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field methods*, *29*(1), pp.3-22.

Haji Sismat, M.A.B., 2016. *Quality and productivity: A comparative analysis of human translation and post-editing with Malay learners of Arabic and English* (Doctoral dissertation, University of Leeds).

Hamilton, S. and Chervany, N.L., 1981. Evaluating information system effectiveness-Part I: Comparing evaluation approaches. *MIS quarterly*, pp.55-69.

Hammond, M. and Wellington, J., 2012. *Research methods: The key concepts*. Routledge.

Hatim, B. and Mason, I., 2014. *Discourse and the Translator*. Routledge.

Hayes, A. (2020). *'Z-Score', Corporate Finance & Accounting.* Available at: https://www.investopedia.com/terms/z/zscore.asp (Accessed: 11 December 2020)

He, P., 2014. *The application of computer-assisted translation tools to the teaching of scientific and technological translation English to Chinese* (Doctoral dissertation, Imperial College of London).

Homiedan, A., 2001. *muqadima fy altarjma al'aal*iya (introduction to MT). Al Obaikan.

House, J., 2014. Translation quality assessment: Past and present. In *Translation: A multidisciplinary approach* (pp. 241-264). Palgrave Macmillan, London.

Hu, K. and Cadwell, P., 2016. A comparative study of post-editing guidelines. *Baltic Journal of Modern Computing*, *4*(2), pp.346-353.

Hughes, M.A. and Hayhoe, G.F., 2009. *A research primer for technical communication: Methods, exemplars, and analyses*. Routledge.

Int, A., 2005. Translator training and modern market demands. *Perspectives: Studies in translatology*, *13*(2), pp.132-142.

Ivanova, O., 2016. Translation and ICT competence in the globalized world. *Procedia-Social and Behavioral Sciences*, *231*, pp.129-134.

Ive, J., 2017. *Towards a Better Human-Machine Collaboration in Statistical Translation: Example of Systematic Medical Reviews* (Doctoral dissertation, Université Paris-Saclay (ComUE)).

Izwaini, S., 2006. Problems of Arabic machine translation: evaluation of three systems. *The British Computer Society (BSC), London*, pp.118-148.

Jabak, O., 2019. Assessment of Arabic-English translation produced by Google Translate. *International Journal of Linguistics, Literature and Translation (IJLLT) ISSN*, pp.2617-0299.

Jia, Y., Carl, M. and Wang, X., 2019. How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, *31*, pp.60-86.

Kalchbrenner, N. and Blunsom, P., 2013, October. Recurrent continuous translation models. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1700-1709).

Kaufman, R. and Keller, J.M., 1994. Levels of evaluation: beyond Kirkpatrick. *Human Resource Development Quarterly*, *5*(4), pp.371-380.

Kenny, D. and Doherty, S., 2014. Statistical machine translation in the translation curriculum: overcoming obstacles and empowering translators. *The Interpreter and translator trainer*, *8*(2), pp.276-294.

Kiraly, D., 2000. *A Social Constructivism Approach to Translator Education*. Jerome Publishing.

Kingscott, G., 2002. Technical translation and related disciplines. *Perspectives: Studies in translatology*, *10*(4), pp.247-255.

Kiraly, D. and Hofmann, S., 2016. Towards a postpositivist curriculum development model for translator education. *Towards authentic experiential learning in translator education*, pp.67-87.

Kirkpatrick, J.D. and Kirkpatrick, W.K., 2016. *Kirkpatrick's four levels of training evaluation*. Association for Talent Development.

Koehn, P., 2009. A process study of computer-aided translation. *Machine Translation*, *23*(4), pp.241-263.

Koponen, M., 2015, November. How to teach machine translation post-editing? Experiences from a post-editing course. In *4th Workshop on Post-Editing Technology and Practice (WPTP4)* (Vol. 2). Miami: Florida.

Koponen, M., 2016a. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, *25*, pp.131-148.

Koponen, M., 2016b. *Machine translation post-editing and effort: Empirical studies on the post-editing process* (Doctoral thesis, University of Helsinki).

Krings, H.P., 2001. *Repairing texts: empirical investigations of machine translation post-editing processes* (Vol. 5). Kent State University Press.

Lange, R.T., 2011. Inter-rater reliability. *Encyclopedia of Clinical Neuropsychology*, p.1348.

Lee, J. and Liao, P., 2011. A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation & Translation Review*, *4*(2).

Loffler-Laurian, A.M., 1996. *La traduction automatique*. Presses Univ. Septentrion.

Lam, T.C. and Bengo, P., 2003. A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, *24*(1), pp.65-80.

Malmkjær, K. ed., 2004. *Translation in undergraduate degree programmes* (Vol. 59). John Benjamins Publishing.

Mellinger, C.D., 2017. Translators and machine translation: knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, *11*(4), pp.280-293.

Miller, R.L. and Brewer, J.D. eds., 2003. *The AZ of social research: A dictionary of key social science research concepts*. Sage.

Moorkens, J., 2018. What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, *12*(4), pp.375-387.

Moorkens, J. and O'Brien, S., 2013, September. User attitudes to the post-editing interface. In *Proceedings of MT Summit XIV workshop on post-editing technology and practice* (pp. 19-25).

Moorkens, J. and O'Brien, S., 2015, May. Post-editing evaluations: Trade-offs between novice and professional participants. In Proceedings of the 18th annual conference of the European association for machine translation (pp. 75-81).

Patton, M.Q., 2002. *Qualitative research & evaluation methods*. sage.

Munday, J., 2016. *Introducing translation studies: Theories and applications*. Routledge.

Nastasi, B.K., Hitchcock, J., Sarkar, S., Burkholder, G., Varjas, K. and Jayasena, A., 2007. Mixed methods in intervention research: Theory to adaptation. *Journal of Mixed Methods Research*, *1*(2), pp.164-182.

Nida, E.A., 1964. *Toward a science of translation*. Brill.

Niño, A., 2008. Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, *21*(1), pp.29-49.

Nitzke, J. and Oster, K., 2016. Comparing translation and post-editing: An annotation schema for activity units. In *New directions in empirical translation process research* (pp. 293-308). Springer.

Noesgaard, S.S. and Ørngreen, R., 2015. The Effectiveness of E-Learning: An Explorative and Integrative Review of the Definitions, Methodologies and Factors That Promote e-Learning Effectiveness. *Electronic Journal of E-learning*, *13*(4), pp.277-289.

O'Brien, S., 2002. Teaching post-editing: a proposal for course content. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*.

O'Brien, S. and Moorkens, J., 2014. Towards intelligent post-editing interfaces. In W. Baur, B. Eichner, S. Kalina, N. Kessler, F. Mayer, & J. Orsted (Eds.), Proceedings of the 20th FIT world congress (pp. 131–137).

O'Brien, S., 2006. *Machine-translatability and post-editing effort: An empirical study using Translog and Choice Network Analysis* (Doctoral dissertation, Dublin City University).

O'Brien, S., 2012. Translation as human–computer interaction. *Translation Spaces*, *1*(1), pp.101-122.

Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L. and Zoran, A.G., 2009. A qualitative framework for collecting and analyzing data in focus group research. *International journal of qualitative methods*, *8*(3), pp.1-21.

Patton, M.Q., 2002. *Qualitative research & evaluation methods*. sage.

Phakiti, A., 2015. Quantitative research and analysis. *Research methods in applied linguistics: A practical resource*, pp.27-48.

Phillips, P.P., 2003. *Training evaluation in the public sector*. The University of Southern Mississippi.

Piotrowska, M. and Tyupa, S., 2014. Translation pedagogy: a new sub-discipline of translation studies. *inTRAlinea*.

Plitt, M. and Masselot, F., 2010. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *Prague Bull. Math. Linguistics*, *93*, pp.7-16.

Pontus, P.S., Matheson, G.J., Schiffler, B.C. and Thompson, W.H., 2017. The readability of scientific texts is decreasing over time. *eLife*, *6*.

Pym, A., 2017. *Exploring translation theories*. Routledge.

Randolph, J.J., 2008. *Multidisciplinary methods in educational technology research and development*. HAMK Press/Justus Randolph.

Reio, T.G., Rocco, T.S., Smith, D.H. and Chang, E., 2017. A critique of Kirkpatrick's evaluation model. *New Horizons in Adult Education and Human Resource Development*, *29*(2), pp.35-53.

Reinke, U., 2018. State of the art in translation memory technology. *Language technologies for a multilingual Europe*, *4*, p.55.

Reiß, K. and Vermeer, H.J., 1984. *Grundleging einer allgemeinen Translationstheorie, Tübingen: Niemeyer of Difference*. Routledge.

Richards, G., Yeoh, W., Chong, A.Y.L. and Popovič, A., 2019. Business intelligence effectiveness and corporate performance management: an empirical analysis. *Journal of Computer Information Systems*, *59*(2), pp.188-196.

Rico Pérez, C. and Torrejón, E., 2012. Skills and Profile of the New Role of the Translator as MT Post-editor. *Tradumàtica*, (10), pp.0166-178.

Robinson, O.C., 2021. Conducting thematic analysis on brief texts: The structured tabular approach. *Qualitative Psychology*.

Rossi, C., 2017. Introducing statistical machine translation in translator training: from uses and perceptions to course design, and back again. *Revista Tradumàtica: tecnologies de la traducció*, (15), p.48.

Şahin, M., 2014, January. Using MT post-editing for translator training. Session 6-Teaching around MT. In *Tralogy II. Trouver le sens: où sont nos manques et nos besoins respectifs?* (p. 14p).

Saldanha, G. and O'Brien, S., 2014. *Research methodologies in translation studies*. Routledge.

Saraireh, M.A., 2001. Inconsistency in technical terminology: A problem for standardization in Arabic. *Babel*, *47*(1), pp.10-21.

Saunders, M.N., Lewis, P., Thornhill, A. and Bristow, A., 2015. *Understanding research philosophy and approaches to theory development.*

Savory, T., 1952. *The Art of Translation*. London.

Schäfer, F., 2003. MT post-editing: how to shed light on the "unknown task". Experiences at SAP. In *EAMT Workshop: Improving MT through other language technology tools: resources and tools for building MT*.

Scherer, R., Siddiq, F. and Tondeur, J., 2019. The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, pp.13-35.

Senez, D., 1998. The machine translation Help desk and the Post-editing Service. *Terminologie et Traduction*, *1*, pp.289-295.

Sieny, M.E.I., 1989. Machine translation in Saudi Arabia. In *Proceedings of Machine Translation Summit II*.

Smidt, A., Balandin, S., Sigafoos, J. and Reed, V.A., 2009. The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability*, *34*(3), pp.266-274.

Sosoni, M.S.V., 2020. MT output and post-editing effort: Insights from a comparative analysis of SMT and NMT output and implications for training. *Fit-For-Market Translator and Interpreter Training in a Digital Age*, p.151.

Soudi, A., Zbib, R., Neumann, G. and Farghaly, A., 2012. Challenges for Arabic machine translation. *Challenges for Arabic Machine Translation*, pp.1-165.

Stebbins, R.A., 2001. *Exploratory research in the social sciences* (Vol. 48). Sage.

Stolze, R., 2011. *The translator's approach: Introduction to translational hermeneutics: theory and examples from practice* (Vol. 41). Frank & Timme GmbH.

Strunk, K.S., 2000. *Status of and barriers to financial impact evaluations in employer-sponsored training programs* (Doctoral dissertation, University of Arkansas)

Sullivan, L.M., Weinberg, J. and Keaney Jr, J.F., 2016. Common statistical pitfalls in basic science research. *Journal of the American Heart Association*, *5*(10), p.e004142.

Suojanen, T., Koskinen, K. and Tuominen, T., 2014. User-centered translation. Routledge.

Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Sycz-Opoń, J. and Gałuskina, K., 2017. Machine Translation in the Hands of Trainee Translators–an Empirical Study. *Studies in Logic, Grammar and Rhetoric*, *49*(1), pp.195-212.

Tatsumi, M., 2010. *Post-editing machine translated text in a commercial setting: Observation and statistical analysis* (Doctoral dissertation, Dublin City University).

TAUS (2013) *Adequacy/Fluency Guidelines*. Available at: <https://taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines />. (Accessed: 15 June 2019)

TAUS Videos (2015) *TAUS Post-Editing Webinar for Arabic module.* 15 September. Available at: https://www.youtube.com/watch?v=i9ZOMRLLHiU (Accessed: 11 April 2018).

Teddlie, C. and Tashakkori, A., 2006. A general typology of research designs featuring mixed methods. *Research in the Schools*, *13*(1), pp.12-28.

Teixeira, C., 2011. Knowledge of provenance and its effects on translation performance in an integrated TM/MT environment. In *Proceedings of the 8th international NLPCS workshop–Special theme: Human-machine interaction in translation. Copenhagen Studies in Language* (Vol. 41, pp. 107-118).

Temizöz, Ö., 2012. Machine translation and postediting. *The European Society for Translation Studies Research Committee State-of-the-Art Research Report*.

Thawabteh, M.A. and Territories, O.P., 2013. The intricacies of translation memory tools: With particular reference to Arabic-English translation. *The International Journal of Localization*, *12*(1), pp.79-90.

Thomas, E.V., Wells, R., Baumann, S.D., Graybill, E., Roach, A., Truscott, S.D., Crenshaw, M. and Crimmins, D., 2019. Comparing traditional versus retrospective pre-/post-assessment in an interdisciplinary leadership training program. *Maternal and child health journal*, *23*(2), pp.191-200.

Tirkkonen-Condit, S., 1990. A Think-Aloud Protocol Study. In *Learning, Keeping, and Using Language: Selected Papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987* (Vol. 1, p. 381). John Benjamins Publishing Company.

Unluer, S., 2012. Being an insider researcher while conducting case study research. *Qualitative Report*, *17*, p.58.

Uszkoreit, H. and Lommel, A., 2013. Multidimensional quality metrics: a new unified paradigm for human and machine translation quality assessment. *Localisation World, London*, pp.12-14.

Varela Salinas, M.J., 2007. How new technologies improve translation pedagogy. *Translation Journal. 11*(4) [online]. Available from: http://translationjournal.net/journal/42technology.htm [Accessed: 21 April 2020].

Salinas, M.J.V., 2020. Methods of Instruction and the Use of New Technologies in Translator Training. *번역· 언어· 기술*, *1*, pp.119-155.

Vermeer, H.J., 1989. Skopos and Translation Commission. *The Skopos and the Commission or Translation Brief*, 29.

Vermeer, H. J. (1989). Skopos and Translation Commission. The Skopos and the Commission or Translation Brief, 29.

Vieira, L.N., 2016. *Cognitive effort in post-editing of machine translation: evidence from eye movements, subjective ratings, and think-aloud protocols* (Doctoral dissertation, Newcastle University).

White, R.M., 1993. Machine Translation Technology: A Potential Key to the Information Age. In *Report of the FCCSET Committee on Industry and Technology, PB-93-134336*. Office of Science and Technology Policy Washington, DC.

Whitehead, D. and Whitehead, L., 2016. Sampling data and data collection in qualitative research. *Nursing and Midwifery Research*, *5*, pp.111-123.

Yang, Y., Wang, X. and Yuan, Q., 2020. Measuring the usability of machine translation in the classroom context. *Translation and Interpreting Studies*. *qualitative report*, *20*(2), pp.134-152.

Yin, R.K., 2003. Case study research: design and methods (ed.). *Applied social research methods series*, *5*.

Yin, R.K., 2018. *Case study research and applications: design and methods*. Sage.

Yngve, V.H., 2000. EARLY RESEARCH AT MI. T. IN SEARCH OF ADEQUATE THEORY. *Early years in machine translation: memoirs and biographies of pioneers*, *97*, p.39.

Young, J. and Kallemeyn, L., 2019. Testing the Retrospective Pre-test with High School Youth in Out-of-School Time Programs. *Journal of Youth Development*, *14*(1), pp.216-229.

Young, J. and Wessnitzer, J., 2016. Descriptive statistics, graphs, and visualisation. In *Modern statistical methods for HCI* (pp. 37-56). Springer.

Zaghouani, W., Habash, N., Obeid, O., Mohit, B., Bouamor, H. and Oflazer, K., 2016, May. Building an Arabic machine translation post-edited corpus: Guidelines and annotation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1869-1876).\

كلية اللغات والترجمة (2019) الرؤية والرسالة والأهداف. [ONLINE] Available at: https://colt.ksu.edu.sa/ar/vission-translation-studies-dept. (Accessed: 5 May 2019