British Journal of Education
Vol. 4, No. 11, pp. 24-34, October 2016
Published by European Centre for Research Training and Development UK (www.ejournals.org)

A REVIEW OF EDUCATIONAL ASSESSMENT: RELIABILITY, VALIDITY AND RELATIONSHIP WITH LEARNING—THE CASE OF NORTHERN IRELAND (NI) TRANSFER TESTS POLICY AND PRACTICE

Md Shidur Rahman
A Doctoral Student at School of Social Sciences, Education and Social Work, Queen's University Belfast, UK

ABSTRACT: *Northern Ireland 11^+ transfer tests policy is a long-standing debatable issue. Presently, the transfer-tests are divided into two distinct test types and they are colloquially known as the AQE (the Association of Quality Education) as well as the GL (Granada Learning) tests which are non-statutory as the government removed the NI transfer tests in 2008. But, previously these tests were called 11^+ exams in which all students took the same tests for grammar school admission. This study aims to evaluate the current NI transfer test policy in light of its reliability, validity, and relationship with learning. The analysis of NI transfer tests traces a number of complications and dilemmas such as unfaithful scoring and grading systems as they contain a lack of transparency. The tests policy also fosters a conflict between the sense of deprivation and advantage. The policy also bewilders a group of pupils, and develops some negative effects on learning. In a word, there are little positive outcomes of these testing systems. Rather, a serious disastrous effect has been culminated in the absence of government care. Henceforth, an alternative transfer testing procedure is essential to be embedded in the NI education system which can fit well with all students in general.*

KEYWORDS: validity, reliability, transfer test policy, learning

INTRODUCTION

Assessment is an influential aspect in education (Taras, 2008) though it is challenging in a contemporary society (McDowell, 2010). Assessment is regarded as 'of learning', 'for learning' and 'as learning' (Black & Wiliam, 1998, 2009; Hume & Coll, 2009). Its use is obvious in various forms, systems, and purposes at different levels and disciplines of education. For example, an assessment process—termed as transfer test—exists in Northern Ireland (NI) education system. Children take this test at the age of 11(Machin, McNally & Wyness, 2013). The test is used in order to select the pupils who have ability to study in grammar schools (Gallagher, 2015). This study wishes to evaluate the current NI transfer test policy in light of its reliability, validity and relationship with learning. The reasons lie in choosing to investigate these three imperative facets of the NI transfer test policy are firstly that most of the researchers, such as Gallagher and Smith (2000); Gardner and Cowan (2000, 2005); and Cowan (2007), studied validity and reliability focusing on the context of old 11^+ transfer test. However, a few researchers, e.g. Elwood (2013), explored the current transfer tests' validity and reliability. Even, Elwood (2013) chiefly concentrated on the validity from an ethics

ISSN 2054-6351 (print), ISSN 2054-636X (online)

viewpoint. Secondly, no study on the present NI transfer test system is so far conducted in combination with validity, reliability and relationship with leaning. Therefore, there exists a vacuum of research which is necessitated to address. Lastly, the study will yield an understanding of the problems with the test system and possible solutions. The study falls into two sections. The first section introduces the transfer test policy and its context; and the second section presents the evaluation of the test system touching on reliability, validity, and links and relationships with learning.

TRANSFER TEST POLICY AND ITS CONTEXT

Context

As already mentioned, this study focused on the transfer tests procedure (as an assessment policy) in the context of NI education system. So, prior to presenting the transfer test policy, there is a need to highlight the NI education system that has a multiple features—particularly the primary, secondary and grammar school systems are segregated on the basis of religion, sex, age, and pupils' ability to be selected for the grammar schools. Then, there are Catholic and Protestant schools. Along with these religious schools, there are some integrated schools in which pupils of any religious background can have access (Gallagher, Smith & Montgomery, 2003) and Irish medium schools. The primary schools cater for pupils from Year 1 to Year 7, the secondary schools from Year 8 to Year 12—although a minority takes pupils up to Year 14 (Gallagher & Smith, 2000), and the Grammar schools from Year 8 to Year 14 (Birrell & Heenan, 2013; Gallagher & Smith, 2000). The pupils from primary level are actually selected for the grammar school level with an assessment system known as transfer test.

Transfer Test Policy

The present NI transfer test seems to be a vexed issue. It was previously known as '11 plus exam' in which all pupils took the same test in order to get admitted into the grammar schools. But, the government eliminated this test system in 2008 with a view to establishing a comprehensive education which was described as a "one size fits all education system" (BBC, 2015; Elevenplusexams, 2015). However, there was a protest against the government decision, and a new transfer procedure emerged in the NI education system. Two different tests have been devised in the new system—the AQE (Association of Quality Education) exams utilised by state/protestant schools; and the GL (Granada Learning) assessment exams used by Catholic schools. The AQE test resembles the old 11^+ test though the science test is excluded. But, the GL test is considerably different from the old 11^+ system, particularly, with respect to test construct and marking strategy. None of these testes are regulated by the government. As a result, these tests are known as unofficial or unregulated AQE and GL test (Lloyd, 2013). In the AQE test, children sit three tests and the best two tests' scores, out of three, are aggregated, while children take two papers in GL test. The AQE test costs £42, and marked by the experienced markers; by contrast, the GL test is free, and machine-marked. The pupils in AQE test write their answers on the test booklets; conversely, pupils in GL test store their answers on a mark-sheet. Then, the AQE test

ISSN 2054-6351 (print), ISSN 2054-636X (online)

follows the scoring system (based on the number) for the result; on the other hand, the GL test adheres to the grading system (based on letters such as A, B1, B2, C1, C2, and D). The children wishing to go to a Protestant school sit the AQE test, and those who desire to go to a Catholic school sit the GL test; on the contrary, children can also sit the both tests as some schools, though few, accept one and the other tests (AQE, 2015; AQUINAS, 2015). The analytical discussion up to this point clarifies that the current transfer test system is full of complications and confusions. Hence, there is a need to evaluate the tests for further understanding.

EVALUATION OF NORTHERN IRELAND (NI) TRANSFER TEST POLICY AND PRACTICE

The term 'assessment' needs to be defined as it is central to this study. Taras (2005, p. 467) said that, "Assessment refers to a judgement which can be justified according to specific weighted set goals, yielding either comparative or numerical ratings." Then Stobart (2008, p. 1) argued, "Assessment, in the form of tests and examinations, is a powerful activity which shapes how societies, groups and individuals understand themselves." Gipps (1994, p.vii) glossed that assessment entails—, "a wide range of methods for evaluating pupil performance and attainment including formal testing and examinations, practical and oral assessment, classroom based assessment carried out by teachers and portfolios." Having taken all these definitions into account, it seems that assessment incorporates a multiple perspectives: judgement, various methods such as tests, examinations, practical and oral assessment and so on. Sebatane (1998 cited in Medland, 2014) described assessment as an overarching concept that incorporates almost every prospect of education. Similarly, Elwood and Lundy (2010, p. 335) stated that, "Assessment is a powerful umbrella term that incorporates a diverse range of actions and process."

In order to elucidate the definition further, types and purposes of assessment are in need of investigation. First, assessment is of different types. These are summative assessment (Assessment of learning): for example, GCSE, A-LEVEL, high-stakes test, national exam; formative assessment (Assessment for learning): classroom based-assessment; and so on (Gipps, 1994; McDowell, Wakelin, Montgomery & King, 2011; Black, 1998). Second, assessment may have various purposes such as to support the learning, to report the achievements of individuals, and to satisfy the demands for public accountability (Black, 1998). Rowntree (1987) mentioned six reasons of assessment from Brian Klug's (1974) thirty-two reasons for formal assessment. These are selection by assessment, maintaining standards, motivation of students, feedback to students, feedback to the teacher, and preparation for life. The selection by assessment—out of these six—is considerably pertinent to the NI Transfer Tests, because the tests chiefly select those pupils who are capable to study in grammar schools. Some also attempted to designate the tests as achievement tests or selection tests. For example, Lloyd, Devine and Robinson (2011) mentioned the test as the selection procedure and as the 11^{+} test.

However, many argued that it is a high-stakes test. Gardner and Cowan (2000), for instance, classified the test as high-stakes because of its serious consequences to the

ISSN 2054-6351 (print), ISSN 2054-636X (online)

pupils who do not get a place at grammar school. Gardner and Cowan (2000) are certainly right in characterising the test as high-stakes because the test results are employed to qualify whether a pupil can get a place or not, and this decision to qualify pupils is consequential due to its influences on students, teachers, schools, communities, and so on (Madaus, 1988). Moreover, this test is likely to be considered, from a wider viewpoint, as a summative assessment. That is to say, it is not a classroom based test because it is not wholly related to what children do in the classrooms. The study will next consider the reliability and validity of NI transfer test procedure.

Reliability and Validity of NI Transfer Test Policy

Reliability

The words such as score, mark, grade, and result are chiefly focused on when defining and measuring reliability of assessment. William (1992, p.1) used the word 'results' while defining reliability as, "an assessment procedure would be reliable to the extent that two identical students would get the same assessment results"; and Feldt and Brennan (1989, p.106) claimed that, "It is almost impossible to deal with issues of definition, quantification, and estimation of reliability without addressing the concept of true score." It seems from these definitions that reliability is basically about marking or score or grade. This reliability issue merits further elaboration as other factors with score or grade—such as errors in marking, variations in grading, inappropriate interpretation of test results and scores, and wrong disclosure and fidelity of assessment—can decline the reliability of assessment.

As in Northern Ireland transfer tests procedure, errors surrounding results are evident, for example, thirty-four candidates received wrong results in 2014 (BBC, 2014), and another student was given D grade unexpectedly with wrong marking in 2010 (Paddyq, 2010). Ricketts (2010) called it a false negative (those deemed to have failed when they were actually qualified). This kind of error may generate negative public perceptions about the test reliability, and seriously undermine their trust in the test system. QCA (2003 cited in Gardner, 2013) took strong position against the error and declared that grading error of any sort, for both the individual student and the system, is unacceptable. Many writers, however, have challenged the QCA's claim on the grounds that assessment inaccuracy is inevitable—it is seldom possible to entirely eliminate the error. No set of results ever be reliable (Newton, 2005a, 2005b), all assessment systems are subject to error (Ricketts, 2010). Newton's and Ricketts's claims ring true when Gardner (2013) indicated that educational assessment is a probabilistic process; however, as the NI transfer test is a high takes test—so, errors in scoring are not desirable. Haladyna and Downing (2004) stressed that a stronger assurance of score accuracy is required in high stakes testing. Consequently, the test agencies (AQE and GL) should be aware of the public's unawareness of assessment inaccuracy, and should expose the strengths and weaknesses of the assessment. In addition to this, the agencies should take the culpability for the error and increase the transparency of grading system in order that the public could trust the test (Newton, 2005b).

Unlike the overall assessment inaccuracy, some definite issues are necessitated to be addressed. For example, the grading system does not seem to be reliable because the

ISSN 2054-6351 (print), ISSN 2054-636X (online)

problems of grade allocation as well as misclassification and misinterpretation of candidates' grades are the matter of a great concern in NI transfer test procedure (Gardner & Cowan, 2000; Harlen, 2006). Cole and Zieky (2001) stated that the testing data present individual variation, not group variation, and this is the major fairness concern. Similarly, Cowan (2007) pointed out that the lack of technical fidelity makes the test unreliable to the stakeholders: parents and students. That is to say, total score or sub-score, and standard measurement errors or test information functions are not reported to the stakeholders. In a nutshell, the information on the reliability of the test are not made available to the public (Gardner & Cowan, 2000). All these evidence draw an issue of having no trust and transparency in the test grading system, and ultimately the blame goes on to the test agencies: AQE and GL. However, Newton (2005b) contended that tests and examinations are deemed to be blunt, and assessment results are thought of as estimates; therefore, it would be naive to criticise test agency only.

In spite of believing that measurement inaccuracy is an inescapable feature of measurement, there is an emphasis on grade descriptors, marking guides and exemplars in order to increase the assessment transparency and to assist pupils' understanding the requirements and standards of assessment (Rust, Price & O'Donovan, 2003; Handley & Williams, 2011). Otherwise, for example, the lack of certitude around grades may fuel another debate among the public. For enhancing the reliability of the test, the AQE and GL should arrange many more training programmes for the parents and the students to clarify the test procedure. The study of Bell, Mladenovic and Price (2013) found that developing students' understanding of assessment practice means enhancing their learning.

In addition, the above measurement errors, as threats to providing true or reliable score to students (which is argued earlier in this section of study), appear from different sources. Firstly, individual variation: one test taker's health, anxiety, motivation level, concentration, forgetfulness, mental efficiency, carelessness, and subjectivity may vary from others. Secondly, situational factors such as working environment of examinees, non-standardised administration, and classroom setting can narrow the reliability (Gipps, 1994). Thirdly, there are unbalanced items in the test.

The discussion about reliability so far indicates that reliability appears to be the problematic issue of the test itself and its markers. Regardless the problems, the NI transfer tests procedure leaves an opportunity for parents to dispute the grades and to remark the test paper if they are unhappy with the test results (AQE, 2015a).

Validity

Another most important and debated concept in educational measurement is validity (Goldstein, 2015) in which an assessment or a testing instrument is considered fit, to what degree, for the purpose. According to Cole and Zieky (2001), "Validity is not dichotomous; it is a matter of degree." Stobart (2001) presented that validity is, "the extent to which a test measures what it purports to measure." This is a conventional definition of validity. However, the NI transfer tests stated what is to be measured. AQE (2015b), for example, specified that, "The Common Entrance Assessment (CEA) has been developed to meet international test standards. It assesses pupils on their

ISSN 2054-6351 (print), ISSN 2054-636X (online)

English and Mathematics ability." This means a student with the score of 112 will perform better in grammar school than a student with the score of 90. Measuring the ability is perceived to be the pupil's performance. Nevertheless, ability is used in the sense of possibility for performance. Wallace (2008) illustrated that a student with having high ability may perform poor, or a student with low ability may perform well. She also argued that one's innate ability is not evident or indeed whether it exists at all till late in one's school career. Therefore, agreeing with Wallace's (2008) view, it could be commented that it is not merely certain that NI transfer tests measure what they purport to measure.

The current transfer tests' ability measurement is found to be affected when we consider the construct validity of the tests. But prior to moving forward to the construct validity, there is a need to look at that a robust debate prevails between content validity (Cureton, 1951) and construct validity (Cronbach & Meehl, 1955). Lissitz and Samuelsen (2007) argued very strongly that validity is about the test contents, not about the test constructs. They claimed that content is relevant to validity as it is internal; whereas constructs are external—so they can be addressed in other test development system. However, Mesick's (1989) suggestion is not to combine different constructs or paradigms into a single measure albeit they are increasingly interlinked. Evidently, in the NI transfer tests, the AQE (2015a) declared that there is an opportunity, for pupils, of sitting for three tests. It is only mandatory to attend two of these tests. Each test is composed with English and Maths (two separate constructs). There are 32 marks for English and 32 marks for Maths. The test, ostensibly, does not estimate the single construct—each paper is a combination of scores in two subject areas. This is to say, test score is treated as a single measure (Gardner & Cowan, 2005). As a consequence, it is difficult to construe the ability with these combining scores, and the AQE did not provide any statement for how they will infer the ability. By contrast, GL assessment assesses two papers (English and Mathematics) separately—the first paper assesses English; the second paper assesses Mathematics (NICCY, 2010). So, the evidence highlights that the transfer tests are not generally valid concerning the test constructs.

However, Stobart (2001) disputed this conventional approach of test validity and claimed that it is a backdated validity concept that considers validity as a property of a test—but validity is not any more regarded as a fixed property of an assessment (Stobart, 2006). Rather, validity is the property of a test score as the use of a test scores needs to be validated, not the test itself (AERA, 2014 in Sireci & Faulkner-Bond, 2015). This claim may not be entirely accepted because Newton (2012) suggested that validity ought to be evaluated by the interpretation of those who use the test as well as those who publish it. This is not to say that validity is not linked to the test scores at all. As such one-third of grammar schools in 2011 admitted the pupils who achieved even the lowest grade (The Belfast Telegraph, 2011, 23).

Recently, the validity has been viewed more broadly. According to Messick (1989, p.19), "For a fully unified view of validity, it must also be recognised that the appropriateness, meaningfulness, and usefulness of score-based inferences depend as well on the social consequences of the testing. Therefore, social values and social consequences cannot be ignored in considerations of validity." This new approach to validity has given a new understanding of consequential validity.

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Many subscribe to Messick's (1989: 19) unitary position to validity of this kind—for instance: Elwood and Murphy (2002, P. 395):, "The social consequences of assessment also impact substantially on the validity of assessment"; Linn (1997); Shaperd (1997): construct validity; crook, et al, (1996); Gipps (1994); and William (1993). By contrast, Popham (1997) and Mehrens (1997) differed with Messick's framework by arguing that social consequences cannot be amalgamated with validity issues. In addition, Borsboom, Mellenbergh and van Heerden (2004) opposed Messick's (1989) position stating that the unified validity concept is not needed as we believe there is nothing to unify. All this leads to comment that the researchers are profoundly divided, and the validity paradigms are seen to have shifted from one to other. But it should be kept in mind that validity is a fluid and relative issue. Any unfamiliar point in question may arise in a new test situation. Gorin (2007) contended that making validation and validity is a continuous process – "Validity is not a box to be checked yes or no."

Messick (1989) is true to say that social values and consequences should not be ignored in respect with test validity. If we take a look back at the transfer test policy section in this study, some equality and ethical issues of NI transfer test may come out which have significant social consequences. For example, the AQE test is mostly aligned with the old 11^+ test and the practice booklets are available, so the AQE test takers are familiar with the test procedure, while the GL test procedure is unfamiliar to the pupils. Then, the parents have to pay for the AQE test, but the GL test is free of cost. These sorts of uneven arrangements provoke fairness concerns among people. Furthermore, the study of Elwood (2013) highlighted some potential social impacts, for example, the form of these tests indicates that one group of pupils may be more benefited than others such as boys may do well in GL test rather than their counterpart because boys do better on multiple choice questions. Again in relation to the AQE test, girls may be more advantaged than their counterpart as they are good at long responses. Another important aspect is that the GL test may be easier than the AQE's. Additionally, as is noted both tests are non-statutory, it seems the tests are vulnerable and unstable. Therefore, these situations may form a concept that the tests may not sustain longer. This test procedure also influences the children as well as parents when they observe the lack of transparency about how the decisions are taken; become confused of what test to sit; come accross the difficulties of taking the both tests, and so on. All this evidence suggests that the existing NI test procedure has an adverse effect on the students and other stakeholders because the social and cultural experience the students and teachers bring to the test situation is a part of the tapestry of the assessment tasks and outcomes (Elwood & Murphy, 2015). Gardner and Cowan (2005), for instance, stated that the students who do not get a place in grammar school, a sense of failure adds them to a personal disappointment. However, it could be argued that the current test system is rather valid comparing to the old 11^+ transfer test because the decision-making of pupils' place allocation in the current test system is conducted based on the two tests scores (yet the GL test is different), but as in the 11^+ transfer test the single test score was used to decide the place. It has been suggested, however, that the predecessor was more stable than the present system, and the children prefer a new common test, not necessarily as the old version (Elwood, 2013).

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Links between reliability and validity

A nexus is noticed between validity and reliability if we take a closer look at the validity definition of Messick (1989:19) that subsumes reliability (test scores and consequences). Furthermore, evidently reliability and validity overlap; assessment result confidence depends upon both validity and reliability; reliability is a part of validity, not a separate issue—it is subsumed into validity (Black & William, 2006; Black, 1998; Storbat, 2008; Feldt & Brennan, 1989; and Wiliam & Black, 1996). All these arguments indicate that many agree that a strong link exists between validity and reliability. On the other hand, others are not aligned with the view that there is a link between validity and reliability, for example, Hogan (2007) argued that a test performance may have reliability—nevertheless it may not be valid or may have limited validity. Reliability is objective – validity is subjective. However, reliability is linked to validity as Messick's (1989) unitary concepts of validity incorporate reliability such as highlighting appropriateness of scores. Beyond this, reliability issue may affect the validity issue for instance publishing wrong grades or the lack of clarity of the test process can hamper the purpose of the test. The NI transfer test procedure is an example of this kind. The NI transfer test agencies such as AQE (2015) just contended that they are maintaining the validity and reliability of the test. But, so far, neither the AQE nor the GL has published any study, as a response to the public, specifying the issues of validity and reliability and outcomes of the tests.

NI transfer test and learning: links and relationships

Having outlined the two key concepts, reliability and validity, of educational testing and assessment—the NI Transfer Tests, this section underscores the links between the tests and learning. Many researchers have conceded that educational assessment has a strong link with learning. Medland (2014), for example, stated that assessment is a key to student learning and achievement as the primary beneficiary of assessment should be students (Hatzipanagos & Rochon, 2010). Dann (2014) expressed that assessment and learning become inextricably intertwine. Like Dann (2014), Brown, Bull & Pendlebury (2013) said, "Assessment is the cash nexus of learning." Furthermore, the categories of assessment at the very outset of this study have showed that the assessment has links with learning such as formative assessment or assessment for learning, and summative assessment or assessment of learning. Among these assessments, classroom based assessments, in other words formative assessments, are more effective (Hargreaves, Earl & Schmidt, 2002) as they prompt student learning (Stiggins, 1991). But, Elwood (2006) argued that formative assessment is confused theoretically and conceptually. In order to prepare the children for the NI Transfer Test, for instance, extra teaching time and preparation are usually provided by the schools, and it was in a greater extent with the old transfer test (11^{+}) though (Gallagher & Smith, 2000). However, since the current test is unofficial, the Department of education has warned the primary schools not to prepare the children (Black, 2015). Moreover, many parents send their children to the coaching centres for extra lessons (Smith, Birthistle & Farrell, 2000). This means the schools and parents drive the students for learning from the classroom in order to cut a good figure in the tests. Besides, during the preparation at school or private coaching

ISSN 2054-6351 (print), ISSN 2054-636X (online)

centre, the children can learn from the feedback given by the teachers as the teacher's feedback to students provides a stronger link with learning (Gipps, 1999).

In looking to the NI Transfer Tests in connection with learning, the tests seem to serve as triggers for learning when the students realise that to be placed in a grammar school gives them a sense of social standing rather than to be placed in a general secondary school (Remedios, Ritchie & Lieberman, 2005). To summarise, the NI transfer tests are deemed to be involved in learning as students are highly engaged with the tests. Black, Harrison, Hodgen, Marshall and Serret (2010) also pointed out that summative assessment should be a positive part of learning process because students' intense involvement in the test procedure can help them to be benefited rather than to be cheated.

Despite the links above, some researchers found that assessment limits the learning opportunities instead of supporting it, for example—High-stakes tests narrow the curricular content to the tested subject; disintegrate subject knowledge into test oriented parts; and drive teaching to be teacher-centred (Au, 2007). High-stakes tests focus on basic skills rather than on extended tasks (Linn, 1993). For the NI transfer tests, the items in English and Mathematics taught by the teachers in respective schools are specific. The teachers employ different techniques to maximise the learning of facts essential for the pupils to secure the place at grammar schools (Johnston & McClune, 2000). That is, the current transfer tests procedure narrow the learning items down from wide to test-specific.

Then, this test procedure has a serious negative effect on learning of students who fail to achieve a place at a grammar school or even those who cannot enter the test or those who opt out of the test. These students are to struggle to build up the self-esteem in the society (Osborne, 2006). Furthermore, the study of Remedios, Ritchie and Lieberman (2005) found that the external pressures such as pressures from the parents can decline the pupils' interests in the subjects. It is also noted that many students lose their interests for learning after passing the transfer test.

Afterwards, the purpose of the tests as presented earlier is to select or to qualify the children for grammar schools. So the purpose indicates that the tests are not concerned with the pupil learning, but with the pupil selection. Accompanying with the purpose, the test structure, too, affects student's learning. As already identified, the NI Transfer Procedure consists of two distinct tests with two separate formats. The AQE is a written test; while the GL is a multiple-choice test. The former is open, though not entirely; but the later is closed. The pupils, who wish to take both tests, are in dilemma to preparing for the tests.

Following the consequences of the tests purpose and structure, the political stalemate surrounding the tests also impinges upon students' learning. Both the tests are administered unofficially, and therein lies anxiety among the teachers and parents who consider the current tests system as chaotic (Birrell & Heenan, 2013). This sort of unrest atmosphere problematizes the student's learning. Lastly, the discussion around the learning, test links and relationships suggests that the NI transfer test policy seems not to sit well alongside learning although many researchers illustrated the links between

ISSN 2054-6351 (print), ISSN 2054-636X (online)

testing and learning. Nevertheless, at the one extreme, the overall discussion so far tells us that the tests do not sit well with the NI education system. At the other extreme, it may be better to have these tests in existence rather than to have no test at all because the political deadlock left the test procedure vacuum. This is not to suggest, however, that the existing NI transfer tests policy is outstanding.

The analysis of NI transfer tests reliability, validity, and relationship with learning renders a number of complications and dilemmas such as unfaithful scoring and grading systems as they contain a lack of transparency. The tests policy also fosters a conflict between the sense of deprivation and advantage. The policy also bewilders a group of pupils, and develops some negative effects on learning. In a word, there are little positive outcomes of this testing system. Rather, a serious disastrous effect has been culminated in the absence of government care. Henceforth, an alternative transfer testing procedure is essential to be embedded in the NI education system which can fit well with all students in general. In order to make it happen, all political parties should come forward to take effective measures for negotiations with parents, educators, and community leaders.

CONCLUSION

This study wished to judge the ongoing NI transfer tests procedure in regard to reliability, validity and relationship with learning. The study has been able to uncover the pros and cons of this tests procedure in relation to the three influential aspects of assessment. A wider range of limitations of the tests policy emerged: obscurity in marking and scoring, and a sense of uncertainty works associated with the tests due to be unregulated by the government. Most important positive side is that the children are engaged in learning process. However, the strength of this study is that it has underlined the problems and attracted the attentions of the stakeholders. It has also located the dangerous effects of the tests process on the learning. So, this study has apprised the concerned authority to construct a new test policy or to amend the existing policy of transfer tests. Concurrently, some limitations of this study are noticeable. Since it is a small-scale research, the issues of three areas have not been possible to focus extensively. Additionally, this is a secondary research. So, further research needs to be undertaken in the future.

ACKNOWLEDGEMENT: A big thank you to Professor Jannette Elwood (School of Social Sciences, Education and Social Work, Queens University Belfast, UK) for her useful support, guidance and feedback

REFERENCES

AQE (2015a). *Frequently Asked Questions*. Retrieved from: https://aqetest.files.wordpress.com/2013/06/gs-15-18-faq-english.pdf.

AQE (2015b). *The Assessments*. Retrieved from: http://aqe.org.uk/the-test/.

AQUINAS (2015). *PPTC FAQs 2015-16: Post Primary Transfer Consortium*. Retrieved from: http://aquinasgrammar.com/wp-content/uploads/2014/05/PPTC-FAQs-2015-16.pdf.

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Au, W. (2007). High-stakes testing and curricular control: A qualitative meta-synthesis. *Educational Researcher*, *36*(5), 258-267

BBC (2015). Transfer tests: Thousands of NI children receive results. The BBC Northern Ireland. Retrieved from: http://www.bbc.co.uk/news/uk-northern-ireland-31068725.

BBC (2014). Rainey Endowed pupils in test marks mix-up get correct results. The BBC Northern Ireland. Retrieved from: http://www.bbc.co.uk/news/uk- northern-ireland-26008237.

Bell, A., Mladenovic, R., & Price, M. (2013). Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, *38*(7), 769-788.

Benett, Y. (1999). *The Validity and Reliabilioty of Assessment and Self-assessments of Work-based Learning*. In Murphy, P. (Ed.), Learners, Learning & Assessment (Pp. 277-289). London: Paul Chapman Publishing Ltd.

Birrell, D., & Heenan, D. (2013). Policy Style and Governing without Consensus: Devolution and Education Policy in Northern Ireland. *Social Policy & Administration*, *47*(7), 765-782. http://dx.doi.org/10.1111/spol.12000

Black, P. & Wiliam, D. (2006). The Reliability of Assessments. In J. Gardner (Ed.), *Assessment and Learning* (pp.119-130). London: SAGE.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31. http://dx.doi.org/10.1007/s11092-008-9068-5

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, *5*(1), 7-74. http://dx.doi.org/10.1080/0969595980050102

Black, P. (1998). *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*, London: Falmer Press.

Black, R. (2015). Transfer test: Warning letters to Northern Ireland schools over coaching sparks new row. The Belfast Telegraph. Retrieved from: http://www.belfasttelegraph.co.uk/news/northern-ireland/transfer-test-warning-letters-to-northern-ireland-schools-over-coaching-sparks-new-row-30949506.html.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215-232.

Borsboom, D., Mellenbergh, G., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111(4),* 1061–1071. http://dx.doi.org/10.1037/0033-295X.111.4.1061

Breslin, G., Brennan, D., Rafferty, R., Gallagher, A. M., & Hanna, D. (2012). The effect of a healthy lifestyle programme on 8–9 year olds from social disadvantage. *Archives of disease in childhood*, archdischild-2011.

Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice,15*(1), 3-17.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, *38*(4), 369-382. http://dx.doi.org/10.1111/j.1745-3984.2001.tb01132.x

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Cowan, P. (2007). Using cat for 11-plus testing in Northern Ireland: what are the issues? IN: Khandia, F. (ed.). *11th CAA International Computer Assisted Conference: Proceedings of the Conference on 10th & 11th July 2007 at Loughborough University*, Loughborough, pp. 129-136.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281. http://dx.doi.org/10.1037/h0040957

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in education*, *3*(3), 265-286. http://dx.doi.org/10.1080/0969594960030302

Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 621 -694). Washington, DC: American Council on Education.

Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 149-166.

Elevenplusexams (2015). 11 Plus in Northern Ireland. Retrieved from: http://www.elevenplusexams.co.uk/schools/regions/northern-ireland-11-plus. Accessed on 07/11/2015.

Elwood, J., & Murphy, P. (2015). Assessment systems as cultural scripts: a Socio-cultural theoretical lens on assessment practice and products. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 182-192. http://dx.doi.org/10.1080/0969594X.2015.1021568

Elwood, J. (2013). Educational assessment policy and practice: a matter of ethics. *Assessment in Education: Principles, Policy & Practice*, *20*(2), 205-220. http://dx.doi.org/10.1080/0969594X.2013.765384

Elwood, J., & Lundy, L. (2010). Revisioning assessment through a children's rights approach: Implications for policy, process and practice. *Research Papers in Education*, *25*(3), 335-353.

Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, *13*(2), 215-232. http://dx.doi.org/10.1080/09695940600708653

Elwood, J., & Murphy, P. (2002). Tests, tiers and achievement: gender and performance at 16 and 14 in England. *European Journal of Education*, *37*(4), 395-416.

Feldt, L. S. and Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement* (pp. 105-146). New York: American Council on Education/Macmillan.

Gallagher, T. (2015). Northern Ireland: An Overview. In Colin Brock (Ed.), *Education in the United Kingdom* (Pp. 255-278). London: Bloomsbury.

Gallagher, T., Smith, A., & Montgomery, A. (2003). *Integrated Education in Northern Ireland*. Participation, Profile and Performance.

Gallagher, T., & Smith, A. (2000). *The Effects Of The Selective System Of Secondary Education In Northern Ireland.* Main Report.

Gardner, J. (2013). The public understanding of error in educational assessment. *Oxford Review of Education*, *39*(1), 72-92. http://dx.doi.org/10.1080/03054985.2012.760290

Gardner, J., & Cowan, P. (2005). The fallibility of high stakes '11-plus' testing in Northern Ireland. *Assessment in Education*, *12*(2), 145-165.

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Gardner, J. and Cowan, P. (2000). *Testing The Test: A Study of the Reliability and Validity of the Northern Ireland Transfer Procedure Test in Enabling the Selection of Pupils for Grammar School Places.* The Queen's University Belfast: Graduate School of Education.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research inEducation*, 24, 355-392. http://dx.doi.org/10.2307/1167274

Gipps, C. (1994). *Beyond Testing: Towards a theory of educational assessment*. New York: Routledge Falmer.

Gipps, C. (1994). Developments in Educational Assessment: what makes a good test? *Assessment in education*, *1*(3), 283-292. http://dx.doi.org/10.1080/0969594940010304

Goldstein, H. (2015). Validity, science and educational measurement. *Assessment in Education: Principles, Policy & Practice*, *22*(2), 193-201. http://dx.doi.org/10.1080/0969594X.2015.1015402

Gorin, J. S. (2007). Reconsidering issues in validity theory. *Educational Researcher*, *36*(8), 456-462. http://dx.doi.org/10.3102/0013189X07311607

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17-27.

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, *36*(1), 95-108. http://dx.doi.org/10.1080/02602930903201669

Harlen, W. (2006). The Role of Assessment in Developing Motivation for learning. In J. Gardner (Ed.), *Assessment and Learning* (Pp.61-80), London: SAGE.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, *39*(1), 69-95. http://dx.doi.org/10.3102/00028312039001069

Hatzipanagos, S. & Rochon, R. (2010). Editorial, *Assessment & Evaluation in Higher Education*, 35:5, 491-492, DOI: 10.1080/02602938.2010.493700.

Hogan, T. P. (2007). *Educational Assessment: A Practical Introduction*. USA: John Wiley & Sons, Inc.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 269-290. http://dx.doi.org/10.1080/09695940903319661

Johnston, J., & McClune, W. (2000). Pupil motivation and attitudes: Self-esteem, locus of control, learning disposition and the impact of selection on teaching and learning. *The effects of the selective system of secondary education in Northern Ireland: Main report (Vol. SEL 5.1). Department of Education in Northern Ireland (DENI).*

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational evaluation and policy analysis*, *15*(1), 1-16. http://dx.doi.org/10.2307/1164248

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice*, *16*(2), 14-16. http://dx.doi.org/10.1111/j.1745-3992.1997.tb00587.x

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational researcher*, *36*(8), 437-448. http://dx.doi.org/10.3102/0013189X07311286

Lloyd, K., Devine, P., & Robinson, G. (2011). Happiest days of our lives. *Belfast: ARK. http://www. ark. ac. uk/publications/updates/update73. pdf. Accessed September*, *5*, 2011.

Lloyd, K. (2013). Happiness and well-being of young carers: extent, nature and correlates of caring among 10 and 11 year old school children. *Journal of Happiness Studies*, *14*(1), 67-80. http://dx.doi.org/10.1007/s10902-011-9316-0

Machin, S., McNally, S., & Wyness, G. (2013). Educational attainment across the UK nations: performance, inequality and evidence. *Educational Research*, *55*(2), 139-164. http://dx.doi.org/10.1080/00131881.2013.801242

Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, *65*(3), 29-46. http://dx.doi.org/10.1080/01619568809538611

McDowell, L. (2010). Challenging assessment?, *Assessment & Evaluation in Higher Education*, 35:3, 263-264, DOI: 10.1080/02602931003690819.

McDowell, L., Wakelin, D., Montgomery, C., & King, S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response. *Assessment & Evaluation in Higher Education*, *36*(7), 749-765. http://dx.doi.org/10.1080/02602938.2010.488792

Medland, E. (2014). Assessment in higher education: drivers, barriers and directions for change in the UK. *Assessment & Evaluation in Higher Education*, (ahead-of-print), 1-16. http://dx.doi.org/10.1080/02602938.2014.982072

Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, *16*(2), 16-18. http://dx.doi.org/10.1111/j.1745-3992.1997.tb00588.x

Messick, S. (1980). Test validity and the ethics of assessment. *American psychologist*, *35*(11), 1012. http://dx.doi.org/10.1037/0003-066X.35.11.1012

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13-103). New York: American Council on Education/Macmillan.

Newton, P. E. (2005a). Threats to the professional understanding of assessment error. *Journal of Education Policy*, *20*(4), 457-483. http://dx.doi.org/10.1080/02680930500132288

Newton, P. E. (2005b). The public understanding of measurement inaccuracy. *British Educational Research Journal*, *31*(4), 419-442. http://dx.doi.org/10.1080/01411920500148648

Newton, P. (2012). *Clarifying the consensus definition of validity*. Cambridge, UK: Cambridge Assessment.

Northern Ireland Commissioner for Children and Young People (NICCY). (2010). Talking transfer. Belfast: NICCY.

Osborne, R. D. (2006). Access to and participation in Higher Education in Northern Ireland. *Higher Education Quarterly*, *60*(4), 333-348. http://dx.doi.org/10.1111/j.1468-2273.2006.00327.x

Paddyq (2010, February 9). Unexpected D result in GL exam [Msg 1]. Message posted to http://www.elevenplusexams.co.uk/forum/11plus/viewtopic.php?t=13503.

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Popham, W. J. (1997). Consequential validity: Right Concern-Wrong Concept. *Educational measurement: Issues and practice*, *16*(2), 9-13. http://dx.doi.org/10.1111/j.1745-3992.1997.tb00586.x

Ricketts, C. (2010). A new look at resits: are they simply a second chance? *Assessment & Evaluation in Higher Education*, *35*(4), 351-356. http://dx.doi.org/10.1080/02602931003763954

Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. Routledge.

Rowntree, D. (1987). *Assessing Students: How shall we know them?* London: Kogan Page.

Remedios, R., Ritchie, K., & Lieberman, D. A. (2005). I used to like it but now I don't: The effect of the transfer test in Northern Ireland on pupils' intrinsic motivation. *British Journal of Educational Psychology*, *75*(3), 435-452. http://dx.doi.org/10.1348/000709904X24771

Rust, C., M. Price, & B. O'Donovan. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, 28(2), 147–64.

Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, *16*(2), 5-24. http://dx.doi.org/10.1111/j.1745-3992.1997.tb00585.x

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting Validity in the Assessment of English Learners. *Review of Research in Education*, *39*(1), 215-252. http://dx.doi.org/10.3102/0091732X14557003

Stiggins, R. J. (1991). Assessment Literacy. *Phi Delta Kappan*, *72*(7), 534-39.

Smith, A., Birthistle, U., & Farrell, S. (2000). Teachers and selection in Northern Ireland. In T. Gallagher, A. Smith (Eds.), *The effects of the selective system of secondary education in Northern Ireland: Main report* (Vol. SEL 6.1). Department of Education in Northern Ireland (DENI).

Stobart, G. (2008). *Testing Times: The uses and abuses of assessment*. London and New York: Routledge.

Stobart, G. (2006). The Validity of Formative Assessment. In J. Gardner (Ed.), *Assessment and Learning* (pp.133-146). London: SAGE.

Stobart, G. (2001). The validity of national curriculum assessment. *British Journal of Educational Studies*, *49*(1), 26-39. http://dx.doi.org/10.1111/1467-8527.t01-1-00161

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466-478. http://dx.doi.org/10.1111/j.1467-8527.2005.00307.x

The Belfast Telegraph (2011, November 23). Even the lowest grade can get you into grammar school: Two out of three took pupils who performed poorly in entrance tests. Retrieved from: http://www.belfasttelegraph.co.uk/news/education/even-the-lowest-grade-can-get-you-into-grammar-school-28683932.html.

Wallace, S. (2008). *Oxford Dictionary of Education.* New York: Oxford University Press.

Wiliam, D. (1992). Some Technical Issue in Assessment: a user's guide. *British Journal of Curriculum and Assessment*, 2 (3), 11-21.

ISSN 2054-6351 (print), ISSN 2054-636X (online)

Wiliam, D. (1993). Validity, dependability and reliability in national curriculum assessment. *The Curriculum Journal*, *4*(3), 335-350. http://dx.doi.org/10.1080/0958517930040303

Wiliam, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*(5), 537-548. http://dx.doi.org/10.1080/0141192960220502

ISSN 2054-6351 (print), ISSN 2054-636X (online)