Standard Assessments: Merits and Demerits and the Alternative Assessments

Narjes Ghafournia[1]

[1] Department of English, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran

Correspondence: Narjes Ghafournia, Department of English, Neyshabur Branch, Islamic Azad University, Neyshabur, Iran. Tel: 98-915-313-0060. E-mail: narjesghafournia@yahoo.com; na.ghafournia@gmail.com

Abstract

Although standardized assessments are extensively applied for major decision making purposes in many language-teaching programs, the tests are not valid and reliable enough for many evaluation programs due to major demerits. Unfortunately, over the years, many high stakes proficiency tests have been widely applied for different educational purposes mainly due to using technically-sophisticated quantitative scoring methods as well as national and international availability. The extensive use of conventional standard tests in different academic settings has resulted in the wide negligence of valid assessment of language leaners' abilities and significant decrease in ethics and fairness in testing. Despite many advantages of standard assessments, some disadvantages of applying standard assessments can be substantially compensated through using formative assessments, criterion- referenced tests, affectively-oriented assessments, culturally familiar tests, and direct assessments such as self-assessment, peer assessment, and portfolio assessment. Combination of different types of assessment is of pedagogical importance due to integration of learning and assessment as well as great emphasis on learners' autonomy and responsibility for second/foreign language learning. The present article discusses the merits and demerits of standardized assessments as well as the necessity of implementing alternative assessments along with standardized assessments in language-learning settings in detail.

Keywords: alternative assessment, self-assessment, peer assessment, portfolio assessment, standard assessment, non-standard assessment

1. Introduction

Standardized tests allow the assessment of test takers' abilities or performances focusing on the comparison of performances of test takers in a norm-referenced way rather than focusing on test takers' performances on their own in a criterion-referenced way. Good examples of standard tests are state educational development tests and college entrance examinations. In comparison, non-standardized tests are commonly used in final project presentations in classes or achievement tests. Essay tests are typically non-standardized whereas some standardized tests such as IELTS and TOEFL tests include writing and oral sections for the assessments of writing and speaking abilities. Other examples of non-standard tests are various types of quizzes or exams, prepared in different question forms, in which multiple responses can be acceptable. Typical examples of non-standard tests are thesis presentations or final project presentations in classes, which no severe comparisons are made between the students' performances in the educational settings concerning particular norms.

The results of standardized tests have a major role in the way states fund educational programs and the way policy makers make major decisions on the curriculum and learning objectives. Standardized tests help instructors account for their students' achievements. Through measuring the degree the students improve in a particular classroom, standardized tests let educational administrators identify the teachers who require further improvements in their instructional approaches. However, many are less likely to make major funding decisions based on non-standardized tests due to lower indexes of reliability and validity of these tests.

In terms of measurement, standardized tests seem more reasonable and acceptable than non-standardized tests, which do not usually provide the opportunity to compare teachers and students through using a particular standard norm. In other words, standardized tests can be usually utilized as diagnostic instruments, detecting problems that learners may encounter in learning contexts. These tests can provide useful norms, depicting the performances of individuals outside and within the normal range. The tests can be applied as part of formal

program evaluation not only at particular schools but also at particular district, regional, or national levels enabling educational authorities to make great decisions on peculiar academic needs of the students of those areas.

However, standardized tests cannot be often used as an evaluation tool in any educational settings, in which learners' needs should be carefully diagnosed. On the other hand, this form of assessment differs from informal non-standard assessment programs, in which particular evaluation and learning needs of learners as well as particular learning objectives and instructional plans are cautiously diagnosed and tailored based on learning contexts. In fact, standardized tests should be used along with non-standard tests in any valid evaluation programs leading to major decisions and further probable adjustments. However, standardized tests are not easily quantifiable. These tests cannot accurately assess some subjects such as creative writing and spontaneous speech that are not easily measurable.

Students who give standardized tests usually have to answer the same or very similar forms of questions. As an example, the reading sections of the TOEFL or IELTS tests have similar questions with the same or similar formats. In comparison, in employing non-standardized tests, test constructors can provide the opportunity for students to answer different types of questions and get involved in accomplishing different types of communicative tasks. As an instance, to carry out a classroom research project, the students may use different types of visual aids to make the content clear. Good examples of common visual aids are using power points, overhead projector slides or transparencies, white or black boards, paper handouts, flip charts, videos, and posters in doing classroom projects.

As non-standardized tests are administered under varying conditions, the way the students' performances are assessed and scored may considerably fluctuate. For example, students may receive higher or lower grades on these tests based on their progress, compared to how they have scored on earlier tests. This considerably affects the reliability of standard tests, which inevitably affects the validity and accountability of these tests.

Constructors of non-standard tests usually take the view that students have different levels of academic achievement and varying degree of relevant background knowledge; therefore, their performances can widely fluctuate. Thus, the majority of students either pass the test or even the most In contrast, in standardized tests, the fluctuation of scores is not so wide struggling students experience great difficulties in passing it.

Standardized tests affect educational standards and opportunities at almost every level, ranging from elementary to college levels to enable teachers to measure their students' achievement and progress in valid and reliable ways. Besides, through using standard testing, moral and social aspects are taken into considerations in the administration and construction of the test. In other words, ethical considerations are taken into consideration by standard test constructors, administrators, and related authorities to interpret test scores far from unwarranted inferences such as invasion of privacy and problem of confidentiality.

The most distinguished features of standard language tests have been outlined by Mousavi (2012) and Brown (2010) as the following:

• Standard assessments presuppose certain standard constant objectives or criteria applied to a broad band of competencies, not usually exclusive or can be removed from particular language-learning curricula. Therefore, standard assessments are designed based on fixed and standard content, which cannot be easily changed from one standard test to another or from one form of the test to another.

• Standard assessments are administered based on pre-determined standard procedures such as time limits, response format, and even the number of desired questions to be answered. Besides, there exist strict standard procedures and rules for scoring, which do not often change in different test-taking settings. In other words, the exact and standard scoring procedures are dictated for each standard assessment, regardless of many other factors about the test takers and test-taking contexts.

• Standard assessments are thoroughly investigated, particularly in an empirical way in terms of validity and reliability. Therefore, based on the intended use of the assessment, it should be demonstrated to be valid and reliable.

• Standard assessments are most often typical of norm-referenced assessments, the goal of which is to place test takers on a continuum across a range of scores; therefore, such assessments differentiate test takers by their relative rankings. This feature of standard assessment is particularly emphasized by Brown (2010).

Good examples of standard assessments are Test of English as a Foreign Language (TOEFL), produced by Educational Testing Service in the United States, or the British counterpart test - International English Language Testing System (IELTS), concerned with University of Cambridge Local Examinations Syndicate (UCLES). In

addition, there exist other popular standard assessments for entering into different educational academic programs such as the Scholastic Aptitude Test (SAT), measuring the abilities predictive of success in academic tasks, The Graduate Record Exam (GRE), necessary for entering into many graduate school programs, (GMAT) the Graduate Management Admission Test, and (LSAT) the Law School Aptitude Test. Most of standard tests are administered in particular academic settings (Richards & Schmidt, 2010).

Although standard assessments are usually administered in multiple-choice format, it does not imply that sticking to this format is a prerequisite characteristic of each standard assessment in spite of false impression of many language teachers. Although multiple-choice formats provide the scorers with great facilitation in objective scoring procedure, which significantly develops consistent rating in terms of inter-rater and intra-rater reliabilities, multiple-choice formats are by no means considered as the absolute formats of standardized language assessment as oral production and writing abilities cannot be assessed in valid way. Therefore, some tests such as (TSE) Test of Spoken Language and (TWE) Test of Written English are administered and assessed in more standardized way by (EST) Educational Testing Service in authentic testing formats, which are much closer to Target Language Use or TLU domain.

Prior to scrutinizing the merits and demerits of standardized-based assessments, it is reasonable to explore the most essential features of these assessments to gain a better understanding into the cognitive process of test taking in language-learning settings, the major focus of this paper.

2. The Merit of Standard Tests

Both standardized and non-standardized assessments have their own merits and demerits. Standardized assessments seek to measure the measurable traits, whereas non-standardized assessments measure students' skills, which are noticeable and to some extent significant, but they cannot be quantified. Both forms of assessment can operate alongside in an educational curriculum.

Among the most potential merits of a standard assessment, Mousavi (2012) emphasized the three essential ones. Standard assessments have higher qualities than ordinary teacher-made tests due to a great amount of expense and time, professionally applied to construct, revise, and administer them. In addition, the great number of professional and experienced testing experts, engaged in the mentioned procedures should not be taken for granted.

The next merit of standard assessments is concerned with releasing language teachers from laborious and time-consuming process of test construction and helping them to devote their time to fundamental instructional activities instead. In addition, in such a standard setting, the teachers try to interpret the scores based on the needs, paradigms, and standards of the educational system, instead of general interpretations without peculiar considerations.

Finally, use of standard assessments improves and facilitates communication among test takers, conveying useful information about individuals to educational professionals. This advantage of standard tests is of great significance because it dispels any mystery concerned with test-taking process. Therefore, prevalent misconceptions about what the tests are designed to do as well as what the scores interpret is further clarified. In addition, through test communication, technical procedures regarding the construction and evaluation of the test is fully reported, and important psychometric properties of the test such as validity and reliability are cautiously reported. Test communication also familiarizes test takers with test-taking procedures, which undoubtedly decreases test anxiety to help them to perform effectively in the test. Finally, through test communication, test takers are provided with informing feedback about their own performance on the test.

Eventually, the next merit of standard assessment, which Brown (2010) emphatically indicated, is the great face validity, which makes standard tests very authoritative-looking instrument.

3. The Demerits of Standard Tests

The demerits of standard assessments are rather revealing, particularly in terms of inappropriate application of standard tests in different educational or even vocational settings. In other words, on some occasions, standardized tests are applied inappropriately in the settings in which they should not be used. However, the feasibility and availability of these tests encourage many teachers to use them instead of employing appropriate diagnostic/prognostic tests. For example, in many local educational systems, language teachers apply standard tests in the place of achievement tests to assess students' language abilities, regardless of the unique goal of each educational setting, which is usually in deep contrast with national/international standard assessments.

Unfortunately, due to the easy access to national and international standardized tests, they are inappropriately employed in different settings; therefore, their validity is not confirmed regarding their intended purposes. In fact,

the use of broad standardized tests in different selection programs, for which they are not intended, is not fair at all. However, the misuse of standardized tests cannot be considered as the demerits of these tests, and should not be mistakenly attributed to the disadvantage or shortcoming of standardized tests. To put it simply, the authorities and teachers who apply standardized tests in inappropriate settings should be blamed for misuse of standard tests not the tests themselves.

The next demerit of standard assessment, which Brown (2010) greatly emphasized, is the potential misunderstanding between direct and indirect assessment of standardized tests. An example of indirect assessment of standard tests is the use of TOEFL test before 1996 to assess the test takers' written and spoken abilities, based on their performances on reading comprehension section. Indirect assessment of extensive reading ability is used in the most recent versions of the TOEFL test through using short paragraphs or short passages of reading comprehension. The use of standardized test for indirect purposes has some advantages and disadvantages, which are clarified through examining the mentioned examples. In the pre-1996 TOEFL Test administration, the great amount of expense for administrating a production-based assessment was reduced through offering comprehension tests. Besides, the findings clarified that the use of short reading passages to assess test takers' extensive reading ability was validated, provided that the use of short reading passages for the intended purpose would be preferable due to ease of administration and limited time. However, lack of any significant relationship between different language traits and constructs makes applying indirect assessment rather invalid.

Therefore, in most of standardized tests, while certification, selection, monitoring, and accountability are legitimate purposes for administration, the role test administrators perform in helping or hindering test takers' learning process should not be neglected. Each type of assessment is part of educational process, which should improve learners' knowledge, understanding, and skills, in the particular intended orientation (Broadfoot, 2005). Therefore, standardized tests should be applied not only in a summative way but also in a formative way, providing insightful information about developing instructional programs in different dimensions. In addition, due to considerable attention given to consequential validly, the impact of standard tests on different educational, psychological, social, and political aspects should not be taken for granted. In fact, through administrating formal standardized summative tests, test takers are usually deprived of the opportunity to convert testing situations into learning experiences to improve their learning process. This does not mean that standardized tests, which are mostly summative, are of no significance, but it means that the integration of formative assessments with standardized testing program either in standardized format or in non-standardized format is of great significance. On the other hand, sometimes, some types of formative assessment can be constructed, administered, and scored in a standardized way, which is of great concern in new trends of language testing paradigms.

Another factor, which has been often taken for granted in many standardized assessment settings, particularly the settings with broader national and international domains, is the emotional aspect of standard assessments. Unfortunately, the influential and crucial role, which emotional factors play in second language learning and testing has been considerably ignored (Csikszentmihalyi et al., 1993). The affective aspect is of particular importance regarding the concept of emotional intelligence in both learning and test-taking settings (Goleman, 1996). For example, if learners like and respect their teacher, feel themselves in a supportive group of peers, accept their classroom culture as quite conducive to learning, and above all, see their own strengths and weaknesses in more reasonable and less competitive way, they are engaged in all classroom activities and act in a more motivated but less competitive way, which undoubtedly leads to effective language learning. Pollard and Filer's work (1996), tracing the learning process of individual students, justified and documented the enormous changes in both learners' effort and achievement when the formal harsh circumstances changed regarding their relationships with teachers, friends, and families. Osborn's et al. (2003) studied the factors influencing the development of a learner's identity in a number of countries. The findings demonstrated clear communalities such as relationships, opportunity to learn and engagement among the country-specific differences, which implies that even in some international standardized tests, emotional factors should be focused and underlined. However, this requires construction and administration of standardized tests in less international broad domain. As an example, consider construction of Asian standardized tests, sharing many cultural and affective considerations, leading to greater consequential validity and reliability. In his comparative study of classroom culture, Alexander (2000) also came to similar conclusions.

From the mentioned points, it is inferred that standardized tests, which are mostly of summative nature are not widely applicable in different educational settings, particularly for decision-making process. They should be either converted into formative assessment or accompanied with other types of formative assessments such as peer-assessment, self-assessment, and portfolio assessment to give the stakeholders and authorities the most

convincing results about the individuals' real positions, which certainly help them make logical and fair decisions about the individuals' real educational and occupational statuses. In other hands, parallel administration of formative and summative assessments is the best alternative.

Due to the importance of parallel administration of standard summative assessment with non-standard formative assessment, a brief elaboration on formative assessment seems quite reasonable. Good examples of non-standard formative assessments are peer-assessment, self- assessment, and portfolio assessment.

4. Formative Assessment

In recent years, due to the growing emphasis on learner's independence and autonomy, self-assessment and peer-assessment have been of great significance. This significance is due to the pedagogic values of these assessments ignored in the conventional trends of language assessment. The degree of integration and interrelation between learning and test- taking processes is so high that some researchers (e.g., Jafarpur, 1991) put great emphasis on the close collaboration between both learners and teachers in controlling the assessment methods, procedures, and outcomes as well as their underlying rationales. The close collaboration between teachers and learners provide ample opportunities for language teachers and course developers to get feedback from students to remove their difficulties and adjust the assessment paradigms and instructional approaches to their actual learning needs.

On the other hand, peer assessment is defined by Topping (1998) as "an arrangement in which individuals consider the amount, level, value, worth, and quality of success of the products or outcomes of learning of peers of similar status" (p.250). The benefits of incorporating peer assessment into the regular formal standardized assessment procedures were discussed in a number of studies (e.g., Burnett & Cavaye, 1980; Earl, 1986; Goldfinch & Raeside, 1990; Kwan & Leung, 1996; Webb, 1995). Peer assessment is of significance because it enables learners to develop abilities and skills in a learning environment, in which they are denied if the teacher assesses their work. In other words, peer assessment provides learners with the opportunity to take responsibility for analyzing, monitoring, and evaluating different aspects of their own learning as well as their peer learning process. In addition, peer assessment can develop students' higher order reasoning and higher level cognitive thought (Birdsong & Sharplin, 1986), help nurture student-centered learning among undergraduate learners (Oldfield & MacAlpine, 1995), encourage active and flexible learning (Entwhistle, 1993), and facilitate a deep approach to learning rather than a surface approach (Entwhistle, 1987, 1993; Gibbs, 1992). Peer assessment can also act as a socializing force, which enhances relevant communication skills and interpersonal relationships between learner groups (Earl, 1986).

Most of the studies in EFL/ESL settings investigated the effect of peer assessment in the area of writing where peers respond and edit each other's written work with the aim of helping each other through revision (e.g., Bell, 1991; Birdsong & Sharplin, 1986; Caulk, 1994; Devenney, 1989; Hogan, 1984; Jacobs, 1989; Jones, 1995; Lynch, 1988; Rainey, 1990; Rothschild & Klingenberg, 1990; Mangelsdorf, 1992; Mendonca & Johnson, 1994; Murau, 1993). Based on the findings of the mentioned studies, through peer assessment, the learners' writing ability, writing performance, and learners' autonomy in the target language learning were significantly developed.

In comparison, there have been fewer studies on the peer assessment of oral presentation skills, reporting significant improvement of language learners in the obtained scores and perceived learning (e.g., Falchikov, 1995, 1995; Mitchell & Bakewell, 1995; Watson, 1989).

There exist further studies, focusing on language learners' feelings or affective reactions toward peer assessment. As an example, in their study, Birdsong and Sharplin (1986) reported that the majority of language learners showed a positive attitude towards evaluating the written work of their peers and being assessed by other peers. Some researchers have taken a more cautious view on the usefulness of peer feedback in writing, arguing for a greater degree of intervention in the process. Therefore, some studies such as Newkirk (1984) and Jacobs (1987) suggested useful ways for language teachers to prepare peers for effective evaluation process through demonstrating how peer evaluation works, explaining peer feedback to the learners, and setting up a structure for groups to work effectively. In contrast, the study of Miller and Ng (1994) showed that the majority of language learners had negative attitudes towards peer assessment, due to some reasons such as subjectivity of the task, unfairness of the whole exercise, strangeness of the activity, and loss of face in front of the classmates. In addition, the learners in the study had negative attitudes toward their peers to evaluate them because they thought that their peers were inexperienced, unqualified, and not linguistically proficient enough to assess them in oral proficiency.

Generally peer-assessment is accompanied with self-assessment in many language-learning and assessment

programs, particularly in learner-centered approaches, in which a collaborative effort exists between teachers and learners in decision-making processes in different aspects of curriculum planning, teaching methodology, and evaluation process (Nunan, 1988). The incorporation of self-assessment into regular formal evaluation program is of great significance because it not only encourages learners and teachers to regard assessment as a shared responsibility, but also opens up wider perspectives on learning process (Huttunen, 1986). Thus, in learner-centered pedagogies, self-assessment plays a central role in shaping and directing the learners' self-autonomy, which prompts the learners to be fully involved in the setting of learning targets and selection of learning activities and materials. Furthermore, the greater importance of self-assessment is likely related to use the language taught in formal instructional context in authentic natural context, which enables the learners to assess their natural autonomous use of the target language. The significance of the incorporation of self-assessment in the formal language evaluation program has been justified by many researchers. Bachman and Palmer (1989, 1996, 2005) considered self-assessment as a valid and reliable measure of communicative language ability. Williams (1992) also reported a close agreement between self-ratings and teacher ratings when teacher ratings are used as actual valid references. Other studies justified a significant high correlation coefficient between teacher-assessment and self-assessment (e.g., Macalpine, 1985; Stefani, 1994; Sullivan & Hall, 1997). On the contrary, some studies recorded a low correlation coefficient between teacher-assessment and self-assessment (e.g., Jafarpur, 1991; Hughes & Large, 1993; Mowl & Pain, 1995; Orsmond et al., 1997).

Despite great significance of self-assessment and peer assessment, one major problem is high subjectivity of these types of assessment, which makes them rather unreliable and hence subject to many criticisms. In fact, this problem does not let the educational authorities use these assessments extensively in many educational programs encouraging them to show greater tendency to apply standardized tests, instead. To develop reliability of non-standard tests, systematic marking criteria is preferred to ensure all markers apply agreed criteria in a consistent fashion, which results in increased validity (Sullivan & Hall, 1997; Woolhouse, 1999).

5. Self-assessment and Portfolio Assessment

Self-assessment is developed as a central feature of portfolio learning and assessment. Portfolio assessment provides a constant link to instruction, and hence it promotes learning reflection and helps learners take responsibility for their own learning. In addition, portfolio assessment enables learners to narrow the gaps in their learning process and enables them to take risks in learning behaviors (Ekbatani, 2000). In other words, portfolio assessment is considered as a means of promoting learner's autonomy. The learners' self-assessment in a target language learning program reflects their learning process. In fact, self-assessment is part of portfolio assessment, in which the learners can be involved in a variety of ways to assess themselves based on the completed portfolios. Portfolio assessment can be accomplished in different ways. For example, sometimes, the learners are required to write an evaluative account of their portfolio experience and submit it together with their portfolio (Hirvela & Pierson, 2000). The learners may rate their portfolio against a checklist of features once used to guide the portfolio process from the beginning. In fact, portfolio assessment can be accomplished in many different ways regarding language components and skills, which are intended to be assessed.

6. Conclusion

In general, it is inferred from this article that standardized assessments should not be solely administered to measure different aspects of language learners' abilities and competencies. Instead, they should be accompanied with different types of non-standard assessments in decision-making purposes in any educational and occupational settings to improve fairness and ethical considerations in test-taking processes.

To put it simply, for selection or placement purposes, administrating standard tests is not enough, but rather it is better to engage language learners in non-standard assessments to select the most qualified and knowledgeable individuals in different settings. In addition, due to great emphasis on the consequential validity of each assessment, formative types of assessments are preferred; therefore, standard assessments should be gradually transformed into formative paradigms of language assessment, apart from their conventional summative nature. In fact, effective standardized assessments should not only be summative but also follow formative paradigms simultaneously.

References

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*(1), 1-34. http://dx.doi.org/10.1207/s15434311laq0201_1

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self ratings of communicative language ability. *Language Testing, 6*, 14-29. http://dx.doi.org/10.1177/026553228900600104

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. New York: Oxford University Press.

Bell, J. (1991). Using peer responses in ESL writing classes. *TESL Canada Journal, 8*, 65-71.

Birdsong, T., & Sharplin, W. (1986). Peer evaluation enhances students' critical judgment. *Highway One, 9*, 23-28.

Broadfoot, P. M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing, 22*(2), 123-141. http://dx.doi.org/10.1191/0265532205lt302oa

Brown, D. A. (2010). *Language assessment principles and classroom practices* (2nd ed.). San Francisco University: Longman.

Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly, 28*, 181-188. http://dx.doi.org/10.2307/3587209

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1993). *Talented teenagers: The roots of success and failure*. Cambridge: Cambridge University Press.

Devenny, R. (1989). How ESL teachers and peers evaluate and respond to student writing. *RELC Journal, 20*, 77-90. http://dx.doi.org/10.1177/003368828902000106

Earl, L. (2003). *Assessment as learning*. Thousand Oaks, CA: Corwin Press.

Ekbatani, G. (2000). Moving toward learner-directed assessment. In G. Ekbatani, & H. Pierson (Eds.), *2000: Learner-directed assessment in ESL* (pp. 1-11). Mahwah, NJ: Erlbaum.

Entwhistle, N. J. (1993). *Recent research on student learning and the learning environment*. Paper presented at the 'New Developments in Learning' Conference, Napier University, Edinburgh.

Falchikov, N. (1995). Improving feedback to and from students. In P. Knight (Ed.), *Assessment for learning in higher education: 1* (pp. 157-166). London: Kogan.

Falchikov, N. (1995). Peer feedback marking: Development peer assessment. *Innovations in Education and Training International, 32*, 175-187. http://dx.doi.org/10.1080/1355800950320212

Gibbs, G. (1992). *Improving the quality of student learning*. Bristol: Technical and Educational Services Ltd.

Goldfinch, J., & Raeside, R. (1990). Development of a peer assessment technique for obtaining individual marks on a group project. *Assessment and Evaluation in Higher Education, 15*, 210-231. http://dx.doi.org/10.1080/0260293900150304

Goleman, D. (1996). *Emotional intelligence: Why it can matter more than IQ*. London: Bloomsbury.

Hirvela, A., & Pierson, H. (2000). Portfolios: Vehicles for authentic self-assessment. In G. Ekbatani, & H. Pierson (Eds.), *2000: Learner directed assessment in ESL* (pp. 105-126). Mahwah, NJ: Erlbaum.

Hogan, P. (1984). Peer editing helps students improve written products. *Highway One, 7*, 51-54.

Hughes, I. E., & Large, B. J. (1993). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education, 18*, 379-385. http://dx.doi.org/10.1080/03075079312331382281

Jacobs, G. (1987). First experiences with peer feedback on compositions: Student and teacher reaction. *System, 15*, 325-333. http://dx.doi.org/10.1016/0346-251X(87)90006-6

Jafarpur, A. (1991). Can naive EFL learners estimate their own proficiency? *Evaluation and Research in Education, 5*, 145-157. http://dx.doi.org/10.1080/09500799109533306

Jones, N. (1995). Business writing, Chinese students and communicative language teaching. *TESOL Journal, 4*, 12-15.

Kwan, K. P., & Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation exercise. *Assessment and Evaluation in Higher Education, 21*, 205-214. http://dx.doi.org/10.1080/0260293960210301

Lynch, T. (1988). Peer evaluation in practice. In A. Brookes, & P. Grundy (Eds.), *Individualization and autonomy in language learning, ELT Documents: 131* (pp. 119-125). London, UK: Modern English Publications.

Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do students think? *ELT Journal, 46*, 158-172. http://dx.doi.org/10.1093/elt/46.3.274

Mendonca, C., & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction.

TESOL Quarterly, 28, 745-769. http://dx.doi.org/10.2307/3587558

Miller, L., & Ng, R. (1994). Peer assessment of oral language proficiency. Perspectives. Working papers of the department of English. *City Polytechnic of Hong Kong, 6*, 41-56.

Mitchell, V. W., & Bakewell, C. (1995). Learning without doing: Enhancing oral presentation skills through peer-review. *Management Learning, 26*, 353-366.

Mousavi, A. (2012). *An encyclopedic dictionary of language testing* (5th ed.). Taiwan: Tung Hua Book Company.

Mowl, G., & Pain, R. (1995). Using self and peer assessment to improve students' essay writing: A case study from Geography. *Innovation in Education and Training International, 32*, 324-335. http://dx.doi.org/10.1080/1355800950320404

Murau, A. (1993). Shared writing: Students' perceptions and attitudes of peer review. *Working Papers in Educational Linguistics, 9*, 71-79.

Newkirk, T. (1984). Direction and misdirection in peer response. *College Composition and Communication, 35*, 301-311. http://dx.doi.org/10.2307/357458

Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524506

Oldfield, K., & MacAlpine, M. (1995). Peer and self-assessment at tertiary level-an experiential report. *Assessment and Evaluation in Higher Education, 20*, 125-132. http://dx.doi.org/10.1080/0260293950200113

Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education, 22*, 357-367. http://dx.doi.org/10.1080/0260293970220401

Pollard, A., & Filer, A. (1996). *The social world of pupil career in primary school*. London: Cassell.

Rainey, K. (1990).Teaching technical writing to non-native speakers. *Technical Writing Teacher, 17*, 131-135.

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). London: Longman.

Rothschild, D., & Klingenberg, F. (1990). Self and peer evaluation of writing in the interactive ESL classroom: An exploratory study. *TESL Canada Journal, 8*, 52-65.

Stefani, L. A. J. (1994). Peer, self, and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*, 69-75. http://dx.doi.org/10.1080/03075079412331382153

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249-276. http://dx.doi.org/10.3102/00346543068003249

Watson, H. M. (1989). *Report on the first year of research into developing an evaluative technique for assessing seminar work*. Collected Original Resources in Education (CORE), Fiche 12 Cl.

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*, 239-261. http://dx.doi.org/10.3102/01623737017002239

Williams, E. (1992). Students' attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education, 17*, 45-58. http://dx.doi.org/10.1080/0260293920170105