

ETHICAL CONFLICTS EXPERIENCED BY IRANIAN EFL TEACHERS IN THE CLASSROOM CONTEXT

Zahra Alimorad

(zahra.alimorad@gmail.com)

Shiraz University

PO Box 1585 Shiraz, Iran 71345

Abstract: Recently, EFL teachers' ethical conflicts in the situated classroom context have got paramount importance. This paper presents findings from an empirical study of ethical conflicts Iranian EFL teachers encounter while engaged in classroom assessment. Critical incidents generated by 49 practicing teachers revealed that a majority of reported conflicts were related to *Do No Harm* principle. Most of the conflicts they encountered involved basic values as one of the conflicting elements while two new conflicting elements emerged in the specific context of this study. It was concluded that Iranian educational policy might need to be changed to mitigate some of these conflicts.

Keywords: classroom assessment, critical incidents technique, ethical conflicts, ethics, Iranian EFL teachers

Testing plays a vital role in any educational system since the decisions that are made on the basis of test scores influence the lives of all of the stakeholders especially those of the test takers. Depending on the scope of the test, different decisions might bring about diverse effects which will be either huge and irreversible or trivial and reversible. To guarantee right decisions to be made on the basis of test scores, test developers try to make their tests as reliable and valid as possible. Messick (1989) as one of the leading figures in this field believes that examining the validity of a given use of test scores is a complex process involving the examination of the evidence that supports and ethically justifies the interpretation or use. However, there seems to be a dearth of research on ethical issues in language testing all over the world. Pointing to the

paucity of research in this area, Bachman (1990, p. 280) maintains that “virtually the only language testing researcher to address the ethical considerations of test use is Spolsky (1981), who has questioned whether language testers have enough evidence to be sure of the decisions made on the basis of test scores.”

Recently, however, the tendency toward more discussions of professionalism and its meaning among language testers has also led to their increased awareness of ethical issues (Bachman, 2000). Some scholars believe that developing ethical standards applicable to a wide variety of assessment contexts is not that straightforward or even possible. For instance, Bachman (2000) states that in order to improve the quality of our professional training, we need standards of professional competence; however, the development of such standards is a challenging endeavor on two grounds, namely, cultural imperialism and ethical relativism. Hamp-Lyons (2000), too, believes that “language testing in different countries, cultures and political contexts will always embody a wide range of variety in practice, and it is hard to see why some methodological practices should be preferred over others on practical grounds” (p. 588) and “what is considered ethically acceptable varies from country to country, culture to culture” (p. 589).

In contrast to the above-mentioned views, some other scholars call for developing ethical standards for all language testing situations. As Davies (1997) convincingly points out, without a set of professional standards, there are no criteria for judging the quality of the tests and the ethicality of their use. Attempting to generate general guiding principles which can be widely applicable to various types of assessment that occur in classrooms, Green, Johnson, Kim, and Pope (2007) developed two such principles, namely, *Do No Harm* (Taylor & Nolen, 2005) and *Avoid Score Pollution*. They defined the former principle as follows:

Educators must be well versed in the potential impact of the practices they use because their assessment and evaluation may have a variety of unintended consequences for their students. For example, a teacher who uses surprise items on a test that did not appear on the study guide may do harm by breaking the implicit bond of trust between teacher and student. Such actions imply lack of respect for student rights and needs. (Green et al., 2007, p.1001)

The latter principle is defined as “any practice that improves test performance without concurrently increasing actual mastery of the content tested.

That is, the score on the test does not represent actual student achievement in the content area and is 'polluted' by factors unrelated to academic attainment (Green et al., 2007, p. 1001). Claiming that test score pollution affects the validity of interpretations and uses that are made of test scores, Haladyna (1992) admitted no test score interpretation or use is valid until we can eliminate the influences of test score pollution. Pennock-Roman (1992, in Haladyna, 1992), on the other hand, claimed that some methods such as test preparation, teaching skills or test wiseness cannot be considered as sources of test score pollution and even Haladyna himself views these methods ethical despite listing them as sources of pollution. She suggested that instead of test score pollution, it is better to use "*test use pollution*" because improper and inappropriate uses of tests by policy makers, school administrators, and teachers are polluting test scores and the interpretations made from them not the tests themselves.

In their study, Green et al. (2007) attempted to define ethical behavior by examining educators' ethical judgments regarding assessment. The study was a web-based survey of educators in which they were asked to read a brief scenario and to indicate whether the student evaluation practice in the depiction was ethical or unethical. Results of their study demonstrated that there was strong agreement among the educators on fewer than half of the scenarios presented in the study. Based on their findings, Green et al. (2007, p. 999) came to the conclusion that "assessment is a realm without professional consensus."

Ethical principles that can guide ethical judgments regarding assessment have also been discussed in some textbooks. Airasian (2005, p. 20), for instance, maintains that the ethical standards for assessment refer to "some aspect of a teacher's fairness in dealing with his or her pupils". However, teachers may not always be well equipped to make ethical judgments related to assessment because of their shaky knowledge base about ethical codes which is the result of either not receiving formal assessment training or receiving out-of-date training (Impara, Plake, & Fager, 1993; Plake & Impara, 1997; Stiggins, 1999).

Additionally, one aspect of ethics in assessment which has rarely been studied is the investigation of ethical conflicts teachers face in their daily classroom assessment practices. While involved in the act of teaching, teachers are always required to make judgments related to assessment. According to Colnerud (1997), the best way to investigate the issues of professional ethics and teaching is by examining the ethical conflicts teachers encounter in their rela-

tionships with the individuals they interact with in their professional life; for example, colleagues, parents, and students.

In his more recent study, Colnerud (2006, p. 366) acknowledges that although theoretical synthesis has been obtained on some approaches to research on teacher ethics, "the difficulty in being a morally good teacher" is in need of further investigation. He believes that this difficulty has arisen from the competing obligations which are present in the teaching profession such as loyalty to colleagues versus acting in the best interests of students. According to him, the organization of the institutions where teachers work may make resolving these conflicts more difficult and complicated.

More recently, Pope, Green, Johnson, and Mitchell (2009) studied ethical conflicts faced by teachers in the United States with regard to the assessment of students. They utilized critical incidents prepared by practicing teachers. Their findings indicated that a majority of reported conflicts were related to score pollution, and conflicts frequently arose between teachers' perceptions of institutional demands and the needs of students. The most frequently reported assessment topics which caused conflict were grading, standardized testing, and special populations. They came to the conclusion that explicit guidelines of assessment practices should be made available for teachers so that they could define and avoid unethical behavior. At the end of their article, Pope et al. (2009) called for more studies regarding ethical conflicts in assessment in different contexts and with different participants so that "perhaps a greater variety of ethical issues would emerge in different populations or different locales" (p. 782). To further this issue, the present study intends to examine ethical conflicts Iranian English as a Foreign Language (EFL) teachers encounter while engaged in their everyday assessment practices. To this end, it seeks answers to the following research questions:

1. Are the ethical categories *Do No Harm* and *Avoid Score Pollution* reported by Green et al. (2007) applicable in the classroom assessment of students in the EFL context of public schools of Iran?
2. What are Iranian EFL teachers' reported ethical conflicts in their daily classroom assessment practices?

METHOD

In Shiraz, public schools are classified and categorized into different districts on the basis of geographical criteria which also roughly correspond to so-

cial and economic ones. To clarify the point, Shiraz city is divided into four educational districts, namely, 1, 2, 3, and 4 among which 1 and 2 are located in the geographically, socially, and economically higher parts of the city while districts 3 and 4 are located in locales where people are of a lower social class and financial status. All four districts are under the control of the Education Organization which is located in Shiraz and which is itself controlled by the Ministry of Education (located in Tehran). The four educational districts of Shiraz were divided into two groups based on their perceived prestige and social class. That is, districts 1 and 2 constituted the first group while districts 3 and 4, the second one. Then, in each group, one district was chosen randomly, that is, district 2 from the first group and district 4 from the second one. Attempt was made to include all of each chosen district's English teachers in the study; therefore, the questionnaire was administered to all of the teachers who were participating in in-service education. The problem, however, was that sometimes a few teachers were absent. Overall, then, 49 in-service English teachers (Male = 21, Female = 28) who have been teaching English in public guidance and high schools in Shiraz took part in the study. Their age ranged from 22 to 58 years with an average of 34.7 years and their teaching experience ranged from 3 to 30 years with an average of 12.8 years. All of them were Iranian and their native language was Persian. The data collected comprised 62 examples of ethical conflicts experienced by these 49 teachers in their assessment practices.

In collecting the data, the researcher used a critical-incident technique similar to Colnerud's (1997) and Pope et al.'s (2009) procedures. Based on this technique, the participants were asked to respond in writing to a single open-ended question about one of their significant experiences in the form of a narrative. The instructions were as follows: "In one or two paragraphs, describe a situation in which you found it too difficult to decide what has been the right or the wrong thing to do from a moral/ethical point of view regarding classroom assessment of students. Be sure that your response will be confidential and your answers will be analyzed anonymously. So, feel free to describe a first-hand experience with ethics and classroom assessment you have encountered during your teaching experience. Then, they were asked to complete, on a separate page, some demographic data such as information about their gender, grade level taught, number of years they have taught, and their age.

After gathering the required data, the researcher and one of her colleagues systematically compared and coded the responses according to their similarities

and differences. Five categories were adapted from Pope et al.'s study (2009), namely, institutional requirements, student needs, teacher needs, parent needs, and basic values and two new categories emerged during the coding process, that is, collegial support and imposed authority demands. According to Pope et al., critical incidents technique allows participants "to respond from their own perspective in their own words and thus better describe the dilemmas they face" (p. 779). Regarding its drawbacks, they refer to its reliance "on the selective and biased nature of the respondent's memory. For example, more recent and vivid events are typically easier to recall" (p.779).

FINDINGS AND DISCUSSION

Findings

Do No Harm vs. Avoid Score Pollution

In order to investigate whether the ethical categories *Do No Harm* and *Avoid Score Pollution* reported by Green et al. (2007) are applicable in the context of Iran, the researcher sorted all the responses into these two categories. Then, she asked one of her colleagues to independently categorize them once more based on the definitions of these terms provided by Green et al. They agreed on labels for 83% of the responses. Differences for the remaining responses were resolved through discussion. Intra-coder reliability was also checked. The researcher herself rated the critical incidents once and then after a few days, she coded them for the second time. The overlap between the two ratings was 89.8% showing agreement on all but eight of the incidents.

Out of all responses coded, 59.3% were classified as consistent with ethical dilemmas centering on "Do No Harm" principle. Pope et al. (2009, p. 779) define this principle in education as "teachers act in such a way as to avoid causing harm to students as well as *other individuals* in schools" (emphasis added). In the particular context of this study, the individuals most vulnerable to harm were the principals of the schools and teachers mostly altered students' test scores for the sake of their benefit. Although in some situations, principals asked the teachers to pass all of the class students regardless of their knowledge and understanding, in most cases, teachers were forced to do so in spite of their own desires and beliefs.

The remainder of responses (40.6%) were classified as “Pollution” of test scores by factors such as consideration of students’ absences or presences in the class, late work, cooperation in classroom activities or lack of it, performance in formative assessments during the term, their physical conditions (e.g. disabled students), their behavior problems and politeness issues, and even their physical appearance (3 cases).

Ethical Conflicts

The researcher examined the content of the incidents more closely and organized them by their conflicting elements. Once more inter- and intra-coder reliabilities were checked. While the index of the former was 83.8% indicating agreement on all but 10 of the conflicts, that of the latter was 91.9% showing agreement on all but five of them. She utilized Pope et al.’s (2009) list of conflicting elements as an organizing framework. As mentioned above, this list consists of five conflicting elements. In addition to those five elements, two other conflicting elements emerged from the incidents in this particular context which were not reported by Pope et al. in the context of their USA study, that is, *Collegial Support* and *Imposed Authority Demands*.

Collegial support is different from “collegial loyalty” introduced by Colnerud (2006, p.378). The former refers to a situation in which one of a teacher’s colleagues who has already done him/her a favor asks him/her to do something in return; for example, to pass a student who is his/her relative. Two teachers reported that they felt terrible to give scores to a student who was recommended by a colleague while his English was very poor. The latter term, however, was defined as “collegial loyalty prevents teachers from defending pupils when other teachers treat them wrongly” (Colnerud, 2006, p. 378).

Imposed authority demands is also different from “institutional requirements” introduced by Pope et al. (2009). They defined the latter term as “rules from the state, district, or school” (p. 781) and included “requirements for standardized test administration, rules for placement in special programs or promotion, exceptions and accommodations for students in special education, or school or district rules about grading practices” (p. 780) in this category while the former term does not refer to district or school rules. In fact, they are against the established rules because on the basis of rules, in Iran, all of the students must be scored objectively and fairly; no additional score must be given to them. In this context, it was the school principals who imposed their au-

thority on the teachers and forced them to give additional unfair scores to low-level students although it was against the established rules. As written by one of the teachers:

During my first years of teaching, once the school principal asked me to change my students' scores and pass all of them. I didn't understand why I should do that. After changing the scores, I was talking about the situation with one of my colleagues. Then the principal quarreled with me because she didn't want anybody to know about it. At that time, I found out that all of us must increase students' scores but nobody must talk about it explicitly.

Table 1 lists the resulting elements of conflict down the left side and across the top along with the frequency of incidents that were classified in each cell. Each incident can be defined by where its two conflicting elements intersect. For example, if a teacher described an incident in which she felt that a student was not competent and knowledgeable enough to pass the test and to go to a higher level but the principal of the school forced her to pass the student otherwise the teacher would be punished in some way, the conflict would fall at the intersection of imposed authority demands and teacher needs. Most of the teachers reported that they had to conform to principals' demands for the sake of their career. They also contended that satisfying principals' desires would lead to more rewarding opportunities. For instance, if they did follow the principal's demands, they would get promotion and rewards for their students' 100% rate of achievement while if they refused to do so, they would be punished or even be fired from that school (because of low rate of their students' success). One teacher wrote that 'it is surprising that students' failure is the teacher's responsibility rather than the students'. But in some other cases where students' failure would not have any serious negative consequences for their teachers, the teachers reported their reluctance to pass those students because they perceived this to be in conflict with their religious or ethical values. Such conflicting situations would fall at the intersection of imposed authority demands and basic values. Table 2 gives a definition of each type of conflict.

After examining the pattern of results, the researcher found that most of the conflicts (61.29%) involved basic values as one of the conflicting elements. By basic values, the researcher meant basic moral/ethical values held by teachers, that is, their beliefs, conscience, and religious considerations. For example, a teacher reported that he passed a student because if he failed the English

course, he would have to leave the school since he failed two other courses as well. In doing that, the teacher felt very confused whether to do it or not because he believed that this student would be in an unfair advantage in comparison to his classmates who had been much better than him. These incidents were considered as a conflict between student needs and basic values. Table 1 also shows that after basic values, 32 (51.61%) of all incidents involved student needs while imposed authority demands as one of the conflicting elements in 18 (29.03%) incidents constituted the third most frequently reported conflicting issue. As the fourth conflicting element, teacher needs constituted 16 (25.80%) cases of the incidents pointing to the fact that teachers in this particular sample were more concerned about their students' problems (32 cases in comparison to 16 ones) rather than their own. Hence, students' needs and desires created more ethical dilemmas for these teachers than their own problems. The fifth conflicting element was institutional requirements with 11 (17.74%) of all incidents reported while parent needs with 7 (11.29%) out of all incidents was the next one and collegial support with 2 (3.22%) of the incidents constituted the last conflicting element.

The present teachers maintained that all of the students would be passed regardless of their knowledge and understanding. Most of the teachers in this study complained about such a situation in which most of the students do not take the classroom seriously, are passive and even make fun of the teachers because they have understood that they would be passed at all costs and nobody would fail at the end of the term.

Discussion

This study intended to examine the ethical conflicts Iranian EFL teachers (who teach English in the public schools) face in their daily classroom teaching experience regarding assessment of students. Considering the first research question posed in this study, it was found that the two ethical categories *Do No Harm* and *Avoid Score Pollution* reported by Green et al. (2007) can also be applicable in the context of Iran. The results, however, indicated that ethical dilemmas centering on *Do No Harm* made up the majority of incidents and this is in contrast to what Green et al. (2007) and Pope et al. (2009) found in their studies. The former study showed that teachers disagreed most about the ethicality of items relating to *Score Pollution* and the latter one revealed also that

ethical dilemmas centering on *Score Pollution* made up the majority of incidents.

With respect to the second research question, this study found that basic values of the teachers are a significant component of ethical dilemmas that teachers face in their classroom assessments. This finding is different from those of Colnerud (1997) and Pope et al. (2009) who found a prevalence of institutional requirements on general ethical dilemmas faced by teachers in Sweden and the United States, respectively. Student needs, imposed authority demands, teacher needs, institutional requirements, parent needs and finally collegial support constituted significant elements of ethical conflicts reported by Iranian public schools teachers, respectively.

Results of this study also showed that Iranian EFL teachers' ethical conflicts are, to some extent, different from those of the participants studied by Pope et al. (2009). In the context of their study, they found institutional requirements as the most conflicting element while in this study, basic values constituted the most conflicting component. In addition, they identified five conflicting elements while seven conflicting components emerged in this study. This discrepancy might point to the fact that ethical issues are context-bound and locally-determined. Thus, the researcher agrees with Hamp-Lyons (2000) who believes that there is a wide range of variety in practice of language testing in different countries, cultures and political contexts and what is preferred in one context might not be preferable in another one on practical grounds. Hence, she is right when claiming "what is considered ethically acceptable varies from country to country, culture to culture" (p. 589).

We must also take into account the fact that ethical issues of schooling "are embedded in an extremely complex and multifaceted context" (Colnerud, 2006, p. 383) and each study can investigate only one phenomenon at a time (Colnerud, 2006). In fact, ethical conflicts in classroom assessment are legion, in part because testing and assessment in general and language assessment in particular, are almost always influenced by complex social and cultural factors that differ widely from one situation to another. Therefore, it might seem too difficult to identify a set of universal criteria applicable to all contexts at all times. Paradoxically, the researcher also agrees with Davies (1997) who points to the need for a set of professional standards for judging the quality of the tests and the ethicality of their use. The development of such standards, however, seems to be a challenging task involving many years of study and investigation on ethical issues in diverse parts of the world.

Table 1. Categorization of Ethical Incidents by Conflicting Elements

Conflicting elements	Institutional requirements	Student needs	Parent needs	Teacher needs	Basic values	Collegial support	Imposed authority demands
Institutional requirements							
Student needs	7						
Parent needs	1						
Teacher needs		3	2				
Basic values	3	22	4				
Collegial support					2		
Imposed authority demands				11	7		

Table 2. Types of Conflicts in Descending Order of Frequency

Types of conflicts	Frequency	Definition/ Explanation
Student needs / Basic values	22	A need or desire of the student conflicts with a basic value held by the teacher.
Imposed authority demands / Teacher needs	11	A need or desire of the principal conflicts with a need or desire of the teacher.
Imposed authority demands / Basic values	7	A need or desire of the principal conflicts with a basic value held by the teacher.
Institutional requirements / Student needs	7	District or school rules conflict with what teacher sees as best for student.
Parent needs / Basic values	4	A need or desire of the parent conflicts with a basic value held by the teacher.
Institutional requirements / Basic values	3	District or school rules conflict with a basic value held by the teacher.
Student needs /Teacher needs	3	A need or desire of the student conflicts with a need or desire of the teacher.
Parent needs / Teacher needs	2	A need or desire of the parent conflicts with a need or desire of the teacher.
Collegial support / Basic values	2	A need or desire of the colleague conflicts with a basic value held by the teacher.
Institutional requirements / Parent needs	1	District or school rules conflict with what parent needs or desires.

CONCLUSIONS AND SUGGESTIONS

Results of the present study confirmed the fact that language testing situations are so unpredictably numerous that we might not be able to postulate a universal set of criteria which help teachers tackle so many unpredictable ethical conflicts they face. Even though, it might, in principle, be possible to devise a set of universal ethical standards, the realities of classroom practice all over the world might pose ethically conflicting situations to teachers which are unique to those situations without being applicable to other contexts. However, more studies in diverse situations with different groups of teachers might help us achieve a unanimous consensus between different points of view with regard to ethical dilemmas.

Most of the teachers in this study considered passing all of the students regardless of their knowledge and understanding a threat to their social and occupational status. Complaining about such a disastrous situation, they admitted their reluctance to pass all of the students because of its negative influences on the educational practices and teacher-student relationships. This is an alarm for those who are responsible for educating students and preparing them for their future life. In fact, such a situation brings about a serious harm to the teachers, their status and their power in the public schools. When students know that teachers have to pass them, they might not respect their teachers and even insult them as mentioned by some of the teachers in this study. Therefore, school principals must be aware of the fact that by trying not to harm themselves, the students and the reputation of their school, they are, in fact, harming the teachers. As mentioned by Green et al. (2007) “Do No Harm” emerges in practice as a choice between harms” (p. 1009) rather than “the avoidance of harm altogether” (p. 1010). It seems that in situations such as those mentioned above, it is better to harm students rather than their teachers, that is, to make students study and try harder rather than to make their teachers pass them.

Therefore, as part of practical implications of the results of this study, it is worth mentioning that evaluating teachers and schools on the basis of students’ success rate seems to be problematic because teachers are only one of a multitude of factors affecting students’ scores. Multiple standard measures are needed to obtain a valid and reliable indicator of their achievement. Therefore, it is suggested that as a possible solution to this problem, the Iranian Education Organization modify its current evaluation criterion of selecting the best schools, the best principals and the best teachers on the basis of the highest rate of their

students' success. For instance, instead of relying on the students' scores and success rates reported by their principals, the Education Organization can administer high-stakes standardized tests to all of the schools on the basis of whose results, the most successful schools can be chosen. In this way, teachers might not be positioned in such ethically conflicting situations where they would like to get promotion and rewards on the one hand, and they want to be fair to all of their students by attempting not to violate the rights of those who are really competent and knowledgeable and who deserve to be passed, on the other.

Trying to find out some of the conditions contributing to the conflicts between imposed authority demands which is an approximately prevalent conflicting element in the context of Iran and teacher needs and basic values, the researcher investigated some of the reasons why the principals made English teachers pass all of the students. The reasons mentioned in the incidents were as follows: in order not to jeopardize the good reputation of the school and its staff, to decrease the school's expenditure by not repeating the exam in September (Shahrivar), to encourage more people to register their children in those schools, and to receive promotion and rewards from the Education Organization because of the high success rate of the students. Additionally, it was found that even though Iranian EFL teachers are aware of and do their best to reconcile ethical conflicts they encounter, rarely are they empowered enough to do so. They need to be supported by powerful authorities in order to be able to challenge principals' demands when they feel that they are unethical and unfair. This is an issue worthy of consideration because as stated by Colnerud (2006, p. 379), "teachers who refer to moral conflicts between taking the pupil into consideration and taking the institution or colleagues into consideration are pointing to a moral phenomenon which is worth taking seriously".

In conclusion, it is worth mentioning that the results of the present study are not definite because of its limitations, the most important of which is its small sample size ($N = 49$), which was chosen only from one city of Iran (Shiraz). Additionally, all of the participants were public school teachers. It is suggested that future research be conducted in other cities of Iran, in private institutes in addition to public schools, and also in other parts of the world in order for the researchers to portray a complete picture of the ethical dilemmas faced by teachers all over the world. However, this study can pave the way for more pieces of research regarding ethics in testing and ethical conflicts in classroom assessment of students.

REFERENCES

- Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17*(1), 1-42.
- Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education, 13*, 627-635.
- Colnerud, G. (2006). Teacher ethics as a research problem: Syntheses achieved and new issues. *Teachers and Teaching: Theory and Practice, 12*(3), 365-385.
- Davies, A. (1997). Introduction: The limits of ethics in language testing. *Language Testing, 14*(3), 235-241.
- Green, S. K., Johnson, R. L., Kim, D. H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education, 23*, 999-1011.
- Haladyna, T. (1992). Test score pollution: Implications for limited English proficient students. *Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on Evaluation and Measurement, 2*, 135-163. Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs.
- Hamp-Lyons, L. (2000). Social, professional, and individual responsibility in language testing. *System, 28*, 579-591.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership, 3*, 510-521.
- Messick, S. A. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (2nd ed., pp. 13-103). Washington: American Council on Education.
- Plake, B., & Impara, J. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom as-*

essment: Learning, achievement, and adjustment (pp. 53-58). San Diego: Academic Press.

Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, *25*, 778–782.

Spolsky, B. (1981). Some ethical questions about language testing. In C. Klein-Braley & D. K. Stevenson (Eds.), *Practice and problems in language testing* (pp. 5-21). Frankfurt, Germany: Peter D. Lang.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, *18*(1), 23–27.

Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson Education, Inc.