Journal of Classroom Interaction, ISSN 0749-4025. © 2014, Vol 49.1, pages 12-25

Using the Dynamic Model to Identify Stages of Teacher Skills in Assessment

Margarita Christoforidou
Cyprus International Institute of Management, Cyprus
Elisavet Xirafidou
National and Kapodistrian University of Athens, Greece

ABSTRACT

The article presents the results of two cross-sectional studies that investigate teachers' skills in using various techniques of assessment in mathematics by taking into account the four phases of assessment. The five dimensions of the dynamic model are also taken into account in proposing a framework for measuring teacher skills in assessment. These two studies were conducted in different countries, Cyprus and Greece, and data were collected through a self-report questionnaire. Semi-structured interviews were also conducted and the internal validity of the study was supported. Using the Rasch and the Saltus models, it was found that assessment skills can be grouped into four types of assessment behavior which are discerned in a distinctive way and move gradually from skills associated with everyday assessment routines to more advanced skills concerned with differentiation in assessment. Comparing the findings of the two studies, it is shown that the same stages were identified through both studies. Implications of these findings for further research are drawn.

INTRODUCTION

Student assessment is considered an integral part of the teaching process. Student assessment is defined as the systematic process of gathering information about student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance (Shepard, 2007). The review of the literature reveals two main purposes of student assessment. First, summative assessment is used for the recording of the overall achievement of a pupil in a systematic way (Mok, 2010). It aims at describing attainment achieved at certain time in order for comparisons to be made according to students' level of performance. Formative assessment is used to identify the strengths and weaknesses of each student (diagnostic aspect) as well as to help teachers plan appropriate next steps in order for improvement to be achieved (intervention aspect). Formative assessment aims at providing information concerning students' performance that could be used for the improvement of both the teaching and learning process (Mok, 2010). Research in the area of educational effectiveness recognizes student assessment as a key effectiveness factor. In particular, teacher effectiveness has been related to the extent that teachers assess their students for formative rather than summative purposes (Creemers & Kyriakides, 2008; Wiliam, Lee, Harrison, & Black, 2004). Similarly, schools with an assessment policy focused on the formative purposes of assessment were

found to be more effective than those giving emphasis to the summative purposes (e.g., De Jong, Westerhof, & Kruiter, 2004; Kyriakides, Campbell, & Gagatsis, 2000).

Although the formative purpose of student assessment has been widely promoted (Popham, 2006; Shepard, 2007) and the need for assessment literate teachers who are able to design and administer more than summative end-of-unit tests (Green & Mantz, 2002) is highlighted; assessment research literature has failed to impact teachers' everyday assessment practice that still appear to be outcome - oriented (Earl & Katz, 2000; Herman, Osmundson, Ayala, Schneider, & Timms, 2006). The various competencies lists developed (e.g., American Federation of Teachers, National Council on Measurement in Education & National Education Association [AFT/NCME/NEA], 1990; National Council of Teachers of Mathematics [NCTM], 1995) describe assessment competencies in relation to general standards of assessment practice without providing details on the specific skills involved. For example, the first standard in the Standards for Teacher Competence in Educational Assessment of Students ([AFT/NCME/NEA], 1990) suggests that teachers should be skilled in choosing assessment methods appropriate for instructional decisions. Although the description of the standard refers to a number of skills a teacher must poses to meet the standard, these skills are not explicitly defined and do not provide a clear picture of what teachers should be able to do. In addition, these lists have not been associated with specific theoretical background and empirical evidence supporting their validity has not been provided. Furthermore, recent conceptions of formative assessment are not addressed (Brookhart, 2011).

Taking the above into consideration, this paper addresses the need for the development and validation of an instrument measuring teacher assessment skills. In the first part of the paper a framework for investigating teachers' skills in student assessment is proposed. Two cross-sectional studies in different countries which investigate teachers' skills in assessment are presented and support of the instrument's validity is provided. Finally, the extent to which the dynamic model can help us generate developmental stages of teacher skills in student assessment is investigated and implications of findings for further research are drawn.

A FRAMEWORK FOR INVESTIGATING TEACHERS' SKILLS IN ASSESSMENT

Recognizing the need for a comprehensive framework based on which skills associated with classroom assessment can be defined and measured, a framework of teacher assessment skills is proposed. The proposed framework takes into account the dynamic nature of assessment and thereby skills associated with each phase of assessment are examined. In addition, assessment skills are defined and measured in relation to teachers' ability to use specific assessment techniques in order to measure different learning outcomes in mathematics. Traditional as well as alternative assessment techniques are taken into consideration, since the literature supports the use of a combination of assessment techniques to assess student learning (Suurtamm, Koch, & Arden, 2010). Moreover, the five dimensions of the dynamic model of educational effectiveness (i.e. frequency, focus, stage, quality and differentiation) which describe the functioning of each effectiveness factor are taken into account.

Given that the use of these five dimensions to measure the assessment factor is described in the previous paper in this issue (see Kyriakides, Archambault, & Janosz, 2013) this section refers to the other two aspects of the framework used to define teacher assessment skills.

A) Main Phases of the Assessment Process.

Classroom assessment is frequently presented in the literature as a cycle subdivided into a number of phases (e.g., Calfee & Masuda, 1997; NCTM, 1995). The most common subdivisions include planning, gathering and interpreting evidence, and use of results. Other important and distinctive aspects of the assessment process discussed in the literature are the construction of assessment tools (De Lange, 1993), assessment administration (Shepard, 2007), recording of assessment information (Kroeger & Cardy, 2006), and communicating assessment results (Stiggins, 2004).

The literature highlights the dynamic relationship between the various phases of the assessment process (Black & Wiliam, 2009). To measure teachers' assessment skills, this study takes into account the four main phases of the assessment cycle (see Figure 1). The division of the assessment process in particular phases is done to make sure that each aspect of assessment practice is taken into account in measuring teacher skills. This division also helps us test the construct validity of the instrument measuring assessment skills. These four phases are based on the assumption that when assessing students effective teachers should make sure that:

a) appropriate assessment instruments are used to collect valid data,
b) appropriate procedures in administering these instruments are followed,
c) data emerged from assessment are analyzed and recorded in an efficient way and without losing important information, and
d) assessment results are reported to parents and students to help them take decisions on how to promote student learning outcomes.

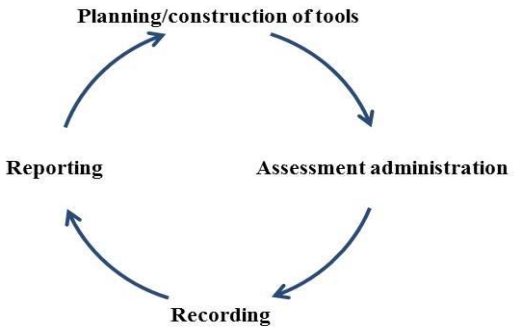


Figure 1. The assessment cycle illustrating the phases of assessment.

Planning and constuction of assessment tools. This phase includes skills involved in the planning and design of assessment, as well as, the construction of the assessment tools. These skills cover decisions concerning the purpose of the assessment, the definition of learning goals against which a student will be assessed, and the development or selection of quality assessment tools through which the purpose and goals of the assessment can be achieved (Pellegrino, Chudowsky, & Glaser, 2001). For example, a teacher is expected to construct assessment tools that are appropriate for the assessment purpose (i.e., formative, summative) she wishes to accomplish.

Administration of assessment instruments. The second phase includes skills associated with the implementation of assessment. These skills refer to decisions concerning the timing of an assessment, assessment's link to instruction, and teachers' role during assessment administration (Shepard, 2007). For example, a teacher is expected to allow appropriate wait time after an oral question and/or give clear instructions in administrating a written test.

Recording and analyzing data. The third phase refers to skills associated with the recording and analysis of data deriving from the assessment process. Documentation allows evidence of performance to be available for future use, interpretation, and revision. It also aids in the identification of gaps in students' learning (Goldhaber & Smith, 2002). Recording assessment information is necessary for information to be effectively used to inform learning and teaching. Skills included in this phase refer to skills associated with the documentation of assessment results (Kroeger & Cardy, 2006) and eliciting information (Schmoker, 2006).

Reporting results to students and parents. The last phase refers to skills related to the communication of assessment results to intended users. The communication of assessment results bridges the gap between the recorded data and their actual interpretation and use by involved participants. In order for intended users to actually act upon assessment information they must first be made aware of such information. Reporting procedures deliver assessment results into the hands of the various intended users of the information in a timely and understandable manner (Roeber, 2003; Stiggins, 2004) and enhance the continuity and quality of students' learning experience (Berry, 2008). Skills included in this phase refer to teacher decisions concerning the purpose of reporting, the audience of reporting, the instruments used to report data, as well as, the quality of teacher communication with parents and students (Stiggins & DuFour, 2009). For example, a teacher is expected to provide students with constructive comments that can help them move their learning forward.

B) Assessment Techniques

Assessment techniques hold an important role in ensuring the quality and effectiveness of assessment since they usually have an influence on how and what students learn (see Kyriakides, 2004). Choosing an assessment technique depends on the target being assessed since student achievement in relation to certain targets can be more appropriately measured by using specific techniques. For example, the assessment of students' skills in oral communication requires the use of different oral assessment techniques rather than the use of written tests. In addition, the use of a variety of techniques allows students to demonstrate different types of learning.

This holds true especially in the case of mathematics since current views of effective mathematic instruction place emphasis on the complexity of mathematics (Boaler, 2008) and require teachers to be able to use a variety of techniques to assess students' conceptual understanding, as well as, their problem-solving and reasoning abilities (Suurtamm et al., 2010). Given the development of alternative assessment methods, as well as, the re-conceptualization of existing traditional methods (Green & Mantz, 2002), it was considered necessary to examine assessment skills in relation to the four most common types of assessment techniques: a) written assessment, b) oral assessment, c) observation and d) performance assessment. For example, it was examined whether different types of written questions were included in teacher tests in order to examine the quality dimension of written assessment. It was also examined how frequently formal and/or informal oral assessment was used to measure student achievement in mathematics. Figure 2 shows the theoretical framework that was used in measuring teacher assessment skills. Specifically, each of the four assessment phases was defined based on the assessment

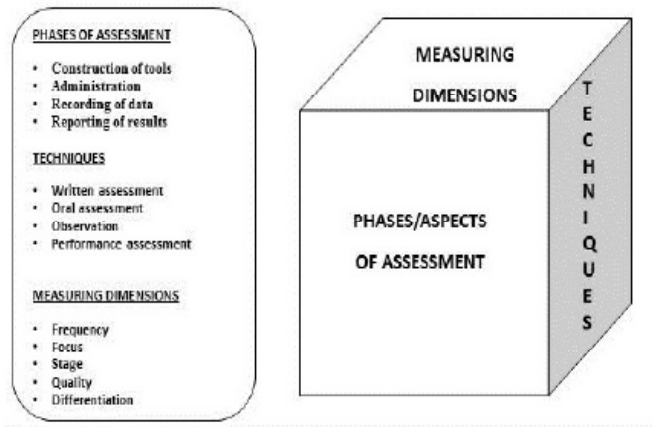


Figure 2. A framework for measuring teacher assessment skills.

skills involved across the five dimensions of the dynamic model and in relation to the use of each of the four most common assessment techniques.

RESEARCH AIMS

This paper has three major aims. The first aim is to develop an instrument based on the framework shown in Figure 2 (p. 15) that could help us measure teacher skills in assessing student knowledge in mathematics. Second, this paper aims to test the validity of the proposed framework while simultaneously examining the construct validity of the instrument that was developed to measure teachers' skills in mathematics assessment. To achieve this aim, data on teacher skills in assessment from two different educational contexts are collected. Finally, drawing from research on teacher developmental theory (Berliner, 1994; Dall' Alba & Sandberg, 2006), this paper aims to examine whether developmental stages can be identified when investigating teachers' skills in assessment. The methodology and findings of the two studies are presented next.

METHODS

As mentioned above, this paper draws on data from two different studies (i.e. Study 1 and Study 2) investigating the validity of the framework used to measure assessment skills. In this section, the participants of each study are presented, information on the research instrument used in both studies is also provided, and issues concerning the testing of the internal validity of each study are addressed. Table 1 (below) presents more details about each study.

Participants

Study 1. Study 1 was conducted in Cyprus at the beginning of school year 2010-2011. Data were collected from a representative sample of 10% of primary education Cypriot teachers. Of the 240 teachers approached, 178 responded which is equivalent to a response rate of 74.2%. Particularly, the research sample consists of 69 men (38.8%) and 109 women (61.2%). Participating teachers' years of experience ranged from 3 to 24 years, with the mean and median of their teaching experience estimated at 10.8 years and 10.5 years, respectively. The teacher

Table 1

Information about the methods used in Study 1 and Study 2

Study	Research Purpose	Location	Participants	Research Instruments	Analysis
Study 1	a) To test the construct validity of the instrument measuring teacher assessment skills. b) To identify stages of teachers' skills in assessment.	Cyprus	178	1. Teacher Questionnaire 2. Semi-structured teacher interviews	Rasch, Saltus, Constant Comparative Method
Study 2	a) To identify stages of teachers' skills in assessment. b) To compare the stages identified in Study 1 with those identified in Study 2.	Attica, Greece	142	1. Teacher Questionnaire 2. Semi-structured teacher interviews	Rasch, Saltus, Constant Comparative Method

sample was found to be representative to the teacher population of Cyprus in terms of gender (X2=0.81, d.f.=1, p=0.42) and years of experience (t=1.21, d.f.=2,578, p=0.22).

Study 2. Study 2 was conducted in Greece at the beginning of school year 2011-2012 in the area of Attica. The fact that the sample of the study comes from only one area of Greece suggests that the results of the study cannot be generalized to the whole population of Greek teachers. Data were collected from 200 primary school teachers. Of the 200 teachers approached 142 responded which is equivalent to a response rate of 71%. The research sample of the second study consists of 41 men (28.9%) and 101 women (71.1%). Participating teachers' years of experience ranged from 1 to 33 years, with the mean of their teaching experience estimated at 18.1 years.

Using Teacher Self-Reports to Measure Assessment Skills

Examining teachers' assessment skills though observations was not possible given the continuous and complex nature of assessment practice. For example, observation of teacher behavior in the classroom may not allow us to measure teacher skills in assessment tool construction or in reporting assessment outcomes since these tasks may take place outside classroom. Thus, the questionnaire was considered as an appropriate tool for measuring a wide range of teacher skills situated at each of the four phases of assessment practice.

Taking into account the three aspects of the proposed framework, a questionnaire consisting of five parts was developed to measure teacher skills in assessment. In the first part, teachers are asked to provide information related to their background characteristics (e.g., gender, position, and years of experience). In the next four parts, teachers are asked to indicate, on a five-point Likert scale, the extent to which they behave in a certain way during Mathematics assessment in their classroom. Each part is concerned with the use of a different assessment technique (i.e., written assessment, oral assessment, observation, performance test). The final part addresses the recording and reporting of assessment results. Each assessment technique is examined in relation to the four aspects of the assessment process (construction, administration, recording and reporting). For each aspect of the assessment process, each of the five dimensions (i.e., frequency, focus, stage, quality and differentiation) is applied. For example, an item measuring the quality dimension of the construction of written assessment asks teachers to indicate whether they include both product and process questions in their written tests, whereas another item inquires whether a specification table is created before developing a written test.

To assess the validity of the proposed framework, as well as the generalizability of the findings of Study 1, Study 2 was conducted in Greece a year after (Xirafidou, 2012). In particular, Study 2 aimed to test whether the teacher questionnaire can be adopted and used in other contexts in order to generate valid data on teacher assessment skills. Since support for the construct validity of the questionnaire was already provided (Kyriakides & Christoforides, 2011), it was decided to conduct the second study in country (i.e. Greece) that shares the same official language and similar educational context in order for the same questionnaire to be used.

Testing the Internal Validity of Each Study

Given that the questionnaire provided self-report data on teacher skills in assessment, issues regarding the validity of these studies could be raised. Therefore, semi-structured interviews were conducted to match responses to different research instruments. In particular, interviews with eight Cypriot teachers in the first study and four Greek teachers in the second study were conducted. The constant comparative method (Maykut & Morehouse, 1994) was used in order to analyze data emerged from all interviews. Initially, `within-case analysis' (Denzin & Lincoln, 1998) of each teacher's responses to the interview was conducted. Then, the results of the qualitative analysis of each interviewee were compared with his/her responses to the questionnaire. Matching teachers' responses from the interviews with the questionnaire data provided support to the internal validity of both studies.

FINDINGS

Teacher' responses to the questionnaire derived from both studies were analyzed by Rasch and Saltus models in order to provide answers regarding the scaling and developmental structure of teachers' abilities in assessment (see Kyriakides et al., 2013). For the sake of brevity, the first part of this section presents the results of Study 1, whereas the second part focuses on the differences that appeared when the same analysis was used to examine the data of Study 2.

A) Scale on Teacher Skills in Assessment and Developmental Stages Emerged from Study 1 Conducted in Cyprus

Since a five-point Likert scale was used to collect data on teacher assessment skills, the extended logistic model of the Rasch model (Andrich, 1998) was applied

Table 2

Statistics relating to the questionnaire measuring assessment skills based on the Rasch analysis of the sample emerged from each study

Statistics	Study 1 Cyprus (n=178)	Study 2 Greece (n=142)
Mean (items*)	0.00	0.00
(persons)	0.08	0.10
Standard deviation (items)	1.12	1.02
(persons)	1.02	0.96
Separability (items)	0.99	0.98
(persons)	0.95	0.94
Mean Infit mean square (items)	0.99	1.00
(persons)	1.00	1.00
Mean Outfit mean square (items)	1.03	1.02
(persons)	1.04	1.07
Infit t (items)	0.04	0.05
(persons)	0.02	-0.04
Outfit t (items)	0.01	0.03
(persons)	0.06	0.05

* L=87 Items

to the sample of 178 teachers and each of the 87 questionnaire items following the same procedure as in the study conducted in Canada reported in the previous issue (Kyriakides et al., 2013). Using the computer program Quest (Adams & Khoo, 1996), it was found that the 87 items of the questionnaire measuring teacher assessment skills had a good fit to the measurement model and that the items were well targeted against the teachers' measures since teachers' scores range from -3.14 to 3.11 logits and item difficulties range from -3.11 to 3.34 logits.

Table 2 provides a summary of the scale statistics for the whole sample of Cypriot and Greek teachers (see columns 2 and 3 respectively). The infit mean squares and the outfit mean squares were found to be near one and the values of the infit t-scores and the outfit t-scores are approximately zero. Moreover, the indices of cases and item separation are higher than 0.92 indicating that the reliability of each scale is satisfactory (Wright, 1985). Thus, the results of the various approaches used to test the fitting of the Rasch model to our data revealed that there was a good fit to the model when teachers' performance to these questionnaire items was analyzed.

Having established the reliability of the scale, the possibility of assessment skills to be grouped into levels of difficulty that may be taken to stand for stages of assessment skills was examined. The various analyses suggested that assessment skills examined can be optimally clustered into four groups of difficulty. Table 3 presents the item difficulty parameters for teachers of the first study. Saltus parameter estimates (i.e., t values) are shown in the bottom of the table. The following observations arise from Table 3. First, item difficulty parameters for teachers in Stage 1 are more spread out than those of the Rasch model, indicating that for these teachers, there is a large gap between the items of Stage 1 and the items in Stages 2, 3, and 4. The gap between the items of Stage 1 and the items of Stage 2 closes considerably when we look at the difficulty estimates that pertain to Stage 2 teachers. Specifically, for teachers who belong to Stage 2, items of Stage 2 are almost as easy as items of Stage 1. As far as the difficulties of items of Stage 3 are concerned, these items are relatively difficult for Stage 2 teachers but for Stage 3 teachers these items are almost as easy as Stage 2 items. Similar observations can be made in relation to items of Stage 4.

Second, a comparison of the segmentation indices reveals that all of them are very large. Similarly, all the asymmetry indices were relatively large. However, the

Table 3

Rasch and Saltus parameter estimates for the 87 items of the teacher questionnaire grouped into four levels of assessment skills

Assessment skills by measurement dimension	Rasch	Stage 1	Stage 2	Stage 3	Stage 4
Frequency: Construction of written tests	-3.11	-3.41	-3.41	-3.41	-3.41
Frequency: Administration of written tests	-3.09	-3.39	-3.39	-3.39	-3.39
Frequency: Recording of homework (check for summative reasons)	-3.07	-3.37	-3.37	-3.37	-3.37
Frequency: Recording of written tests	-3.05	-3.35	-3.35	-3.35	-3.35
Focus: Construction of written tests (various types of written questions)	-3.01	-3.31	-3.31	-3.31	-3.31
Focus: Recording of written tests (single score)	-2.97	-3.27	-3.27	-3.27	-3.27
Focus: Reporting of written tests (parents only)	-2.95	-3.29	-3.29	-3.29	-3.29
Stage: Construction of written tests	-2.93	-3.23	-3.23	-3.23	-3.23
Frequency: Reporting of written tests	-2.91	-3.21	-3.21	-3.21	-3.21
Focus: Construction of written tests (basic skills)	-2.89	-3.19	-3.19	-3.19	-3.19
Stage: Recording of written tests	-2.87	-3.27	-3.27	-3.27	-3.27
Focus: Construction of written tests (product questions)	-2.86	-3.16	3.16	-3.16	-3.16
Quality: recording of written test (for summative reasons)	-2.85	-3.22	-3.22	-3.22	-3.22
Stage: Reporting of written tests (parents only)	-2.84	-3.14	-3.14	-3.14	-3.14
Quality: Construction of written tests (only for summative reasons)	-2.83	-3.13	-3.13	-3.13	-3.13
Quality: Administration of written tests (for summative reasons)	-2.82	-3.21	-3.21	-3.21	-3.21
Quality: Reporting of written tests (for summative reasons)	-2.82	-3.12	-3.12	-3.12	-3.12
Frequency: Administration of oral assessment	-1.10	-1.01	-2.84	-2.80	-2.89
Frequency: Construction of oral assessment (systematic)	-1.09	-1.00	-2.83	-2.79	-2.88
Focus: Construction of written tests (process questions)	-1.08	-0.98	-2.81	-2.77	-2.86
Quality: Construction of written tests (specification table)	-1.07	-0.95	-2.78	-2.74	-2.83
Quality: Administration written tests (clarifications)	-1.05	-0.91	-2.74	-2.70	-2.79
Frequency: Reporting of homework (for measuring basic skills)	-1.03	-0.93	-2.76	-2.72	-2.81
Frequency Reporting of oral assessment	-1.00	-0.90	-2.73	-2.69	-2.78
Quality: Construction of written tests (representative)	-0.98	-0.88	-2.71	-2.67	-2.76
Frequency: Construction of performance test	-0.95	-0.81	-2.64	-2.60	-2.69
Frequency: Administration of performance test	-0.89	-0.79	-2.62	-2.58	-2.67
Stage: Recording written tests (value added)	-0.88	-0.82	-2.65	-2.61	-2.70
Focus: Administration of other forms (homework basic skills)	-0.88	-0.78	-2.61	-2.57	-2.66
Frequency: Recording oral assessment	-0.87	-0.75	-2.58	-2.54	-2.63
Frequency: Reporting observation (non-systematic)	-0,84	-0,74	-2.57	-2,53	-2.62
Stage: Reporting written tests to parents only	-0.83	-0.84	-2.67	-2.63	-2.72
Focus: Construction of performance tests (basic skills only)	-0.81	-0.71	-2.54	-2.50	-2.59
Quality: Construction of written tests (take into account learning needs of students)	-0.78	-0.80	-2.63	-2.59	-2.68
Frequency: Reporting of performance tests	-0.71	-0.65	-2,48	-2.44	-2.53
Stage: Administration of oral assessment	-0.66	-0.63	-2.46	-2.42	-2.51
Focus: Recording of written tests (descriptive comments)	-0.62	-0.58	-2.41	-2.37	-2.46

Table 3 (continued)

Assessment skills by measurement dimension	Rasch	Stage 1	Stage 2	Stage 3	Stage 4
Frequency: Recording of performance tests	-0.59	-0.49	-2.32	-2.28	-2.37
Frequency: Construction of observation	-0.54	-0.44	-2.27	-2.23	-2.32
Focus: Construction of oral tests (basic skills)	-0.52	-0.41	-2.24	-2.20	-2.29
Focus: Administration of oral tests (clarifications)	-0.45	-0.38	-2.21	-2.17	-2.26
Frequency: Administration of observation	-0.38	-0.30	-2.13	-2.09	-2.18
Focus: Construction of observation (basic skills only)	-0.32	-0.27	-2.10	-2.06	-2.15
Focus: Construction of performance tests (basic skills)	-0.30	-0.20	-2.03	-1.99	-2.08
Stage: Construction of oral assessment	-0.27	-0.23	-2.06	-2.02	-2.11
Stage: Construction of performance test	-0.23	-0.18	-2.01	-1.97	-2.06
Stage: Administration of oral assessment	-0.22	-0.12	-1.95	-1.91	-2.00
Stage: Administration of observation	-0.20	-0.10	-1.93	-1.89	-1.98
Stage: Construction of observation	-0.17	-0.04	-1.87	-1.83	-1.92
Focus: Construction of oral assessment (basic skills)	-0.15	-0.05	-1.88	-1.84	-1.93
Stage: Recording of oral assessment	-0.09	-0.01	-1.84	-1.80	-1,89
Stage: Administration of performance test	0.05	0.08	-1.75	-1.71	-1,80
Quality: Reporting of written tests (formative)	0.95	0.99	0,13	-1.66	-1.50
Stage: Administration of observation (systematic and continuous)	0.97	1.09	0.23	-1.56	-1.40
Quality: Construction of observation tools	0.98	1.08	0.22	-1.57	-1.41
Quality: Recording of performance tests (formative reasons)	1.00	1.10	0.24	-1.55	-1.39
Focus: Construction of written tests (complex objectives)	1.03	1.11	0.25	-1.54	-1.38
Focus: Construction of oral assessment (mathematics communication)	1.11	1.00	0.14	-1.65	-1.49
Focus: Construction of performance test (complex objectives)	1.14	1.06	0.20	-1.59	-1.43
Quality: Reporting of performance tests (formative)	1.20	1.07	0.21	-1.58	-1.42
Quality: Recording of oral assessment (multi-dimensional)	1.21	1.11	0.25	-1.54	-1.38
Quality: Recording of performance test (formative – comments on a variety of skills)	1.24	1.17	0.31	-1.48	-1.32
Focus: Administration of observation (complex skills such as communication)	1.25	1.21	0.35	-1.44	-1.28
Focus: Administration oral assessment (maths communication)	1.28	1.19	0.33	-1.46	-1.30
Focus: Administration observation (low-inference observation tools)	1.34	1.21	0.35	-1.44	-1.28
Focus: Construction of observation (focused on specific skills/objectives)	1.35	1.28	0.42	-1.37	-1.21
Focus: Reporting written tests (parents and pupils)	1.37	1.32	0.46	-1.33	-1.17
Focus: Reporting performance tests (parents and pupils)	1.38	1.34	0.48	-1.31	-1.15
Focus: Reporting oral assessment (parents and pupils)	1.42	1.39	0.53	-1.26	-1.10
Focus: Reporting observation (parents and pupils)	1.43	1.33	0.47	-1.32	-1.16
Stage: Construction of observation (systematic and continuous)	1.45	1.43	0.57	-1.22	-1.06

Table 3 (continued)

Assessment skills by measurement dimension	Rasch	Stage 1	Stage 2	Stage 3	Stage 4
Differentiation: Recording written tests	2.60	2.91	2.10	1.98	-0.86
Differentiation: Construction of written tests	2.68	2.96	2.15	2.03	-0.81
Differentiation: Construction of oral assessment	2.78	2.94	2.13	2.01	-0.83
Differentiation: Administration of written tests	2.88	2.97	2.16	2.04	-0.80
Differentiation: Construction of performance tests	2.95	3.00	2.19	2.07	-0.77
Differentiation: Recording performance tests	3.00	3.08	2.27	2.15	-0.69
Differentiation: Administration of performance test	3.08	3.12	2.31	2.19	-0.65
Differentiation: Reporting of written tests	3.12	3.21	2.40	2.28	-0.56
Differentiation: Construction of oral assessment	3.16	3.29	2.48	2.36	-0.48
Differentiation: Administration of oral assessment	3.19	3.38	2.57	2.45	-0.39
Differentiation: Administration of observation	3.23	3.42	2.61	2.49	-0.35
Differentiation: Reporting of performance tests	3.27	3.50	2.69	2.57	-0.27
Differentiation: Reporting of observation	3.29	3.56	2.75	2.63	-0.21
Differentiation: Recording of oral assessment	3.31	3.64	2.83	2.71	-0.13
Differentiation: Recording of observation	3.33	3.69	2.88	2.76	-0.08
Differentiation: Reporting oral assessment	3.34	3.73	2.92	2.80	-0.04

Note 1: Double lines in the body of Table 3 above separate the four stages of assessment as indicated by cluster analysis.
Note 2: The Saltus parameter estimates (i.e. τ values) are shown below in Table 3a.

Table 3a

	Examinee Stage			
Item Class	**1**	**2**	**3**	**4**
1	0.00*	0.00*	0.00*	0.00*
2	0.00*	1.83	1.79	1.88
3	0.00*	0.86	2.65	2.49
4	0.00*	0.81	0.93	3.77

*Fixed at zero for model identification

asymmetry index between Stages 3 and 4 is extremely high. This implies that the transition from one level to the other is not linear and that the transition from Stage 3 to 4 is much more difficult than the transition among the first three levels. Thus, the development of teacher skills in assessment is discontinuous rather than continuous. In addition, the discontinuity in development is much more obvious for teachers moving from Stage 3 to 4. A description of the different levels of teacher behavior in assessment is given below.

Stage 1: Using written tests to measure basic skills in mathematics for summative reasons. The assessment skills included in this stage show that, teachers demonstrating this type of behavior use everyday assessment routines. Stage 1 teachers enrich or alter ready-made written tests and use a variety of types of written questions to assess students' performance. However, they do not use oral assessment and/or observation in a systematic way to assess their students' performance. Records are kept only in relation to written assessment results whereas results are reported to parents for summative purposes.

Stage 2: Using different techniques of assessment to measure basic skills in mathematics. The assessment skills included in this stage reveal that teachers demonstrating this type of behavior are able to use in an appropriate way the various techniques of assessment in order to measure basic skills in mathematics. Specifically, Stage 2 teachers create a specification table before developing their written tests. In this way, they try to ensure that their tests are representative to what has been taught in the classroom. They also include some test items which measure the student ability to give a correct answer to a task and few items which investigate the process that

was used by each student in his/her attempt to find an answer to a problem (i.e., process questions are included). In designing test items, they also take into consideration students' abilities. In addition, they reported that they offer clarification comments to students during assessment administration and that they plan oral assessment and observation. In regard to recording of assessment data, they use descriptive comments to give feedback to their students. Finally, they report to parents on their students' assessment results.

Stage 3: Using assessment techniques to measure more complex educational objectives for formative reasons. Teachers demonstrating this type of behavior are able to use assessment techniques to measure more complex educational objectives in mathematics such as their ability in communicating by using mathematics. Observation is used in a systematic way by setting specific goals and creating observation tools in relation to these goals. Recording is done for data deriving from all assessment techniques and not merely written assessment (as in Stage 2 teachers) and takes the form of goal and/or exercise specific documentation. In addition, reporting is done for formative reasons and assessment information is reported not only to parents but also to their students. Finally, group assessment is systematically used and is concerned with each student's contribution to the team work rather than with the team's overall performance.

Stage 4: Differentiation in assessment: applying assessment in and for different occasions and students. Based on the assessment skills included in this type of behavior it appears that Stage 4 teachers are able to differentiate assessment procedures and tools based on their students' needs. Therefore, teachers of this stage don't use the same assessment tools to measure the achievement of different groups of students and they are more flexible during the administration process (e.g., they give extra tasks to those who finish earlier and more time to slow learners). They also differentiate reporting of assessment information to both parents and students (e.g., reporting is done more often to those needed; they use different forms/language that are in line with the educational level of parents) and pursue teacher-parent communication especially when parents are not visiting the school.

B) Comparing the Results of the two Studies Measuring Teacher Assessment Skills

The same analysis procedures used in Study 1 were used to examine data deriving from Study 2 conducted in Greece. The Rasch model was applied on the whole sample of Greek teachers and all 87 questionnaire items, using the computer program Quest. It was found that all items of the questionnaire measuring teacher assessment skills had a good fit to the measurement model. The third column of Table 2 provides a summary of the scale statistics for the whole sample of Greek teachers. We can observe that data emerged from the whole sample of Greek teachers fit as well to the Rasch model as data emerged from the Cyprus study (see column 2). Moreover, by taking into account the estimates of the Rasch analysis, the assessment skills were grouped into the same four levels of difficulty (i.e., assessment stages) identified through the analysis of data emerged from Study 1 (see Xyrafidou, 2012). However, by comparing the item difficulty estimates emerged from Study 1 and Study 2, six statistically significant differences at level .05 were identified. These are reported in Table 4. Although the differences reported are larger than the standard error of the scale, one can see that none of these

Table 4

Assessment skills with statistically significant differences in their difficulty emerged from the separate Rasch analysis in each country

	Study 1-Cyprus		Study 2-Greece	
Assessment skills	**Rasch**	**Stage**	**Rasch**	**Stage**
Frequency: Reporting of written tests	-2.91	1	-2.67	1
Frequency: Administration of written tests	-3.09	1	-2,82	1
Frequency: Construction of performance test	-0.95	2	-1.09	2
Focus: Construction of observation (focused on specific skills/ objectives)	1.35	3	1.19	3
Differentiation: Construction of oral assessment	2.78	4	2.99	4
Differentiation: Recording performance tests	3.00	4	2.64	4

skills was found to be situated at two different stages. It is also important to note that teachers in each country were found to be situated at a single stage indicating that the same classification of stages can emerge by analyzing either the difficulty level of the items or the Rasch estimates of the persons.

Finally, Figure 3 shows how the Greek and Cypriot teacher sample is distributed in the four stages of assessment skills. We can observe that more than 55% of teachers from each country is situated at the two lower stages. Moreover, in each country, less than 16% of teachers is situated at Stage 4. These findings reveal that there is a need for improving teacher skills in student assessment in each country. Regarding to the distribution of teachers in these four stages, the distribution of the Greek teachers is more normal in shape than the distribution of Cypriot teachers. In Cyprus, the largest percentage of teachers (31.5%) is situated at stage 1 whereas in Greece the largest percentage of teachers was found to be at stage 2 (36.1%).

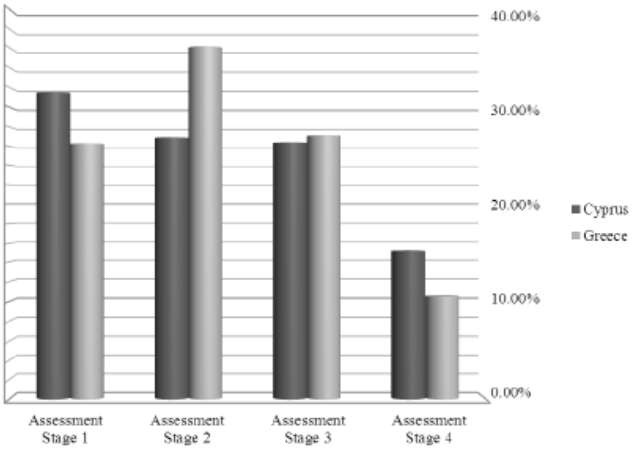
IMPLICATIONS OF FINDINGS

The results of the two studies reported above appear to provide some support to the assumption that the dynamic model can help us generate developmental stages of teacher skills in assessment. The developmental scale proposed was identified in both measuring cases, suggesting the generalizability of the results. The four stages of teacher assessment behavior identified are described in a distinctive way thus addressing a weakness of previous stage related studies to provide a clear picture of what each stage entails (Dall' Alba & Sandberg, 2006). The content of each stage is now specifically determined whereas previous stage models suffered from vagueness and lack of clarity on what could actually constitute each developmental stage. In addition, the four stages identified represent an integrated approach to assessment practice including various functions and purposes of assessment, thus moving away from the commonly applied summative-formative distinction. Looking at the description of the four stages we can see that they move from relatively easy to more advanced types of teacher behavior in assessing student knowledge and skills in mathematics. Starting from skills associated with everyday classroom routines with a mainly summative orientation we can observe a gradual movement towards skills associated with the use of assessment for formative purposes. This is in line with recent literature supporting that effective teachers use formative–oriented assessment in everyday classroom practice (e.g., Creemers & Kyriakides, 2008; Wiliam, Lee, Harrison, & Black, 2004).

Moreover, a specific measurement framework was used to describe not only quantitative but also qualitative characteristics of classroom assessment helping us define and measure specific skills associated with assessment practice. This framework is in line with recent literature and could help clarify the blur area of classroom assessment by directly associating it to specific dimensions. These dimensions enable the measurement of classroom assessment's effectiveness not only in terms of its formative purpose; but also in terms of all aspects of the assessment process. Thus, the framework proposed here could be used to design instruments that examine assessment knowledge and skills in relation to actual assessment practice. In order to support this argument, a teacher questionnaire was designed and data collected from two different countries provided support to the construct validity of the instrument. Moreover, the same stages of assessment were identified by the two studies. Since the studies took place in countries where the same language is used but the features of the educational systems are different, one can see the results reported here as a starting point for conducting further research to test the generalizability of these two studies.

Future studies will have to face more issues in developing comparable questionnaires since not only adaptation but also translation of questionnaire items is needed. However, these two studies show that there is a possibility to identify the same stages of assessment by conducting similar studies in different contexts. To test the generalizability of the findings reported here, we should also investigate whether the developmental stages of assessment skills can also be identified in measuring skills of teachers in assessment of student achievement in various subjects (not only in mathematics) and in

Figure 3. Percentage of Cypriot and Greek teachers situated at each of the four stages of assessment skills.



assessment of students at different phases of schooling (not only at primary school level).

Implications of findings for teacher professional development in assessment can also be drawn. Most attempts to bring about improvement in teachers' assessment practice come from the area of formative assessment and focus on training teachers in the use of assessment strategies recognized as beneficial to students' outcomes (e.g., Black & Wiliam, 2005; Black, McCormick, James, & Pedder, 2006). However until today, there has been no systematic empirical evidence to describe in detail the skills related to effective assessment practice. Investigating which activities associated with the factor of classroom assessment can have positive impact on students outcomes could help us not only define effective assessment practice but also identify appropriate improvement steps.

Given that the aim of both studies was to test the validity of the instrument measuring teacher skills in student assessment as well as the extent to which the dynamic model can help us generate developmental stages of teacher skills in assessment, assessment skill acquisition was not investigated over time. Follow-up studies are needed in order to examine stage identification over a longer period of time. These studies will investigate the extent to which stepwise development of teacher assessment skills can be achieved and if so what type of programs of teacher professional development should be offered in order to improve teacher assessment skills and their effectiveness status.

REFERENCES

Adams, R.J., & Khoo, S. (1996). Quest: The interactive test analysis system, Version 2.1. Melbourne: ACER.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT/NCME/NEA). (1990). Standards for teacher competence in educational assessment of students. Educational Measurement: Issues and Practice, 9(4), 30-32.

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. Applied Measurement in Education, 1(4), 363–378.

Berliner, D. (1994). Expertise: The wonder of exemplary performances. In J. Mangieri & C. Block (Eds.), Creating powerful thinking in teachers and students: Diverse perspectives, (pp. 161–186). Fort Worth, TX: Harcourt Brace College.

Berry, R. (2008). Assessment for Learning. Hong Kong: Hong Kong University Press.

Black, P., & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. The Curriculum Journal, 16(2), 249-261.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31.

Black, P., McCormick, R., James, M., & Pedder, D. (2006). Learning How to Learn and Assessment for Learning: a theoretical inquiry. Research Papers in Education, 21(2), 119-132.

Boaler, J. (2008). What's math got to do with it? London: Penguin Press.

Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30(1), 3–12.

Calfee, R.C., & Masuda, W.V. (1997). Classroom assessment as inquiry. In G. D. Phye (Ed.), Handbook of classroom assessment: learning, adjustment, and achievement (pp. 69-102). San Diego, CA: Academic Press.

Creemers, B.P.M., & Kyriakides, L. (2008). The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools. London: Routledge.

Dall'Alba, G. & Sandberg, J. (2006). Unveiling professional development: a critical review of stage models. Review of Educational Research, 76, 383–412.

De Jong, R., Westerhof, K.J., & Kruiter, J.H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. School Effectiveness and School Improvement, 15(1), 3–31.

De Lange, J. (1993). Assessment in problem-oriented curricula. In N.L. Webb & A.F. Coxford (Eds.), Assessment in the mathematics classroom (NCTM Yearbook) (pp. 197 – 208). Reston, VA: NCTM.

Denzin, N.K., & Lincoln, Y.S. (Eds.) (1998). Collecting and interpreting qualitative materials. Thousand Oaks, California: SAGE.

Earl, L., & Katz, S. (2000). Changing classroom assessment: Teachers' struggles. In N. Bascia & A. Hargreaves (Eds.), The sharp edge of educational change (pp. 97–111). London: Routledge.

Goldhaber, J., & Smith, D. (2002).The development of documentation strategies to support Teacher reflection, inquiry, and collaboration. In V. Fu, A. Stremmel, & L. Hill (Eds.), Teaching and learning: Collaborative exploration of the Reggio Emilia approach, (pp. 147-160). Columbus, OH: Merrill/Prentice-Hall.

Green, S.K., & Mantz, M. (2002). Classroom assessment practices: Examining impact on student learning. Paper presented at the Annual Meeting of the American Educational Research Association (AERA). April, New Orleans, LA.

Herman, J.L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). The Nature and Impact of Teachers' Formative Assessment Practices. National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation Technical Report 703, University of California, Los Angeles, CA (Available online from http://www.cse.ucla.edu/products/reports/R703.pdf).

Kroeger, J., & Cardy, T. (2006). Documentation: A hard-to-reach place. Early Childhood Education Journal, 33(6), 389–98.

Kyriakides, L. (2004). Investigating Validity from Teachers' Perspective through their engagement in Large-Scale Assessment: the Emergent Literacy Baseline Assessment Project. Assessment in Education: Principles, Policy and Practice, 11(2), 143-165.

Kyriakides, L., Archambault, I., & Janosz, M. (2013). Searching for stages of effective teaching: a study testing the validity of the dynamic model in Canada. Journal of Classroom Interaction, 48(2), 11-24.

Kyriakides, L., Campbell, R.J., & Gagatsis, A. (2000). The Significance of the Classroom Effect in Primary Schools: An Application of Creemers' Comprehensive Model of Educational Effectiveness. School Effectiveness and School Improvement, 11(4), 501–529.

Kyriakides, L., & Christoforides, M. (2011). Searching for stages of teacher skills in assessment: Implications for research on teacher professional development. Paper presented at the 37th International Association for Educational Assessment Annual Conference (IAEA) 2011. Manila, Philippines, October 2011.

Maykut, P., & Morehouse, R. (1994). Beginning qualitative research: a philosophical and practical guide. London, UK: The Falmer Press.

Mok, M. M. C. (2010). Self-directed Learning Oriented Assessment: Assessment that Informs Learning & Empowers The Learner. Hong Kong: Pace Publications Ltd.

National Council of Teachers of Mathematics (NCTM). (1995). Assessment standards for school mathematics. Reston, Va.: NCTM.

Pellegrino, J.W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.

Popham, W.J. (2006). Phony formative assessments: Buyer beware! Educational Leadership, 64(3), 86–87.

Roeber, E. (2003). Steps in the right direction: Reporting assessment results to students, parents, school board members, and the media. U.S. Department of Education.

Schmoker, M. (2006). Results now. Alexandria, VA: Association for Supervision and Curriculum Development.

Shepard, L.A. (2007). Formative assessment: Caveat emptor. In C.A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 279–303). Mahwah, NJ: Erlbaum.

Stiggins, R.J. (2004). New assessment beliefs for a new school mission. Phi Delta Kappan, 86(1), 22–28.

Stiggins, R., & DuFour, R. (2009). Maximizing the power of formative assessments. Phi Delta Kappan, 90(9), 640–644.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' emerging assessment practices in mathematics: Classrooms in the context of reform. Assessment in Education: Principles, Policy, and Practice, 17(4), 399-417.

Wiliam, D., Lee, C., Harrison, C., & Black, P.J. (2004). Teachers developing assessment for learning: Impact on student achievement. Assessment in Education: Principles, Policy, and Practice, 11(1), 49-65.

Wright, B.D. (1985). Additivity in psychological measurement. In E.E. Roskam (Ed.), Measurement and personality assessment, (pp.101-112). Amsterdam, the Netherlands: Elsevier Science Publishers BV.

Xirafidou, E. (2012). Investigating Greek primary teachers' skills in mathematics assessment. Unpublished MA thesis, National and Kapodistrian University of Athens, Athens, Greece. [in Greek]

Correspondence regarding this article should be addressed to Margarita Christoforidou at the Cyprus International Institute of Management, Cyprus. Email may be sent to margo.chr@gmail.com