Chapter 17
(Mis)Alignment of Medical Education Validation Research with Contemporary Validity Theory: The Mini-CEX as an Example

Debra (Dallie) Sandilands and Bruno D. Zumbo

Like all educational assessments, assessments of medical students, residents and practicing physicians must be supported by research evidence of their validity for the purposes for which they are used. Evidence for validity is the foundation upon which meaningful and defensible interpretations of assessment results are based. The strongest evidence to support defensible use of an assessment is derived from the alignment of its validation research with contemporary validity theory as described in the *Standards for Educational and Psychological Testing* (the "*Standards*", AERA et al. 1999). The *Standards* provide criteria for the evaluation of all educational and psychological tests, testing practices and the effects of test use, as well as guidelines for test developers and users about sound and ethical use of tests. Sireci and Parker (2006) reviewed court cases involving disputes about educational tests and found that typically it is issues of test validity that are challenged in court, and that testing practices that are closely aligned with the *Standards* are more likely to withstand legal challenge. Thus in high stakes testing environments such as assessment in medical education it seems particularly important to ensure that validation efforts are aligned with contemporary validity theory as expressed in the *Standards*.

Research in other areas such as psychology and general education has found that studies are not providing validity information aligned with contemporary validity theory and that some sources of validity evidence are not being investigated or reported (Cizek et al. 2008, 2010; Hogan and Agnello 2004). Therefore the purpose

D. Sandilands, Ph.D. (✉)
Faculty of Education, Department of Educational and Counseling Psychology, and Special Education, The University of British Columbia, 2125 Main Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: sandilan@mail.ubc.ca

B.D. Zumbo, Ph.D. (✉)
Measurement, Evaluation, and Research Methodology (MERM) Program, Department of Educational and Counseling Psychology, and Special Education, The University of British Columbia, 2125 Main Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: bruno.zumbo@ubc.ca

B.D. Zumbo and E.K.H. Chan (eds.), *Validity and Validation in Social, Behavioral, and Health Sciences*, Social Indicators Research Series 54, DOI 10.1007/978-3-319-07794-9_17,

of this research was to investigate the extent to which studies in medical education are aligned with contemporary validity theory, using the Mini-Clinical Evaluation Exercise (Mini-CEX) (Norcini and Blank 1995) as an example. We investigated studies about the Mini-CEX because it is one of the most extensively used and studied assessment tools in medical education (Kogan et al. 2009). It has been used for more than two decades to evaluate the workplace performance of medical students, residents and physicians. Mini-CEX assessment results may have significant implications for individuals, for the educational programs that train them, and for society that relies on them to provide adequate medical care. Although there is a great deal of research that investigates the Mini-CEX, to date there has been no thorough review of the extent to which the body of Mini-CEX validation research meets the recommendations and criteria set out in the *Standards* or the extent to which the *Standards'* recommended sources of validity evidence are being reported regarding the Mini-CEX.

We conducted a systematic review of Mini-CEX studies to reveal potential gaps or limitations which may guide future Mini-CEX validation research. Specifically, our research questions were:

1. To what extent are validation studies of the Mini-CEX consistent with key aspects of contemporary validity theory as outlined in the *Standards*?; and
2. To what extent have the recommended sources of validity evidence outlined in the *Standards* been reported regarding the Mini-CEX?

It is important to note at the outset that the purpose of this study was not to evaluate the Mini-CEX or the overall quality of the research about the Mini-CEX, nor was our goal to ascertain the degree to which Mini-CEX research supports its use. Rather we were interested in gaining an understanding about how well the research is aligned with current validity theory.

In the following introductory sections we provide an overview of the Mini-CEX and of contemporary validity theory as outlined in the *Standards*.

The Mini-CEX

The Mini-CEX is a direct observation assessment tool originally developed by the American Board of Internal Medicine (ABIM) to assess the clinical skills of internal medicine residents in medical encounters with patients in a broad range of situations and locations (i.e. inpatient, outpatient, or emergency room settings). It was specifically designed to cover the skills most often required by residents in real patient encounters such as medical interviewing, physical examinations, decision-making, counseling, and clinical judgment or reasoning. The Mini-CEX is administered in two parts. First, a faculty member observes a resident while the resident conducts a focused history and physical examination on a patient, and provides a diagnosis and treatment plan. Next, immediately after the patient encounter, the faculty member gives the resident formative feedback both verbally and in writing

on a Mini-CEX rating form. The Mini-CEX rating form is said to be aligned with six (US) Accreditation Council for Graduate Medical Education (ACGME) general competencies, each of which is rated on a scale from 1 to 9. There is one additional rating for "overall clinical competence". Ratings of 1 through 3 reflect unsatisfactory performance, 4 through 6 are satisfactory (but 4 is defined as "marginal"), and 7 through 9 are superior. Each Mini-CEX takes 10–20 min to complete (ABIM 2009).

Since its inception in 1995, the Mini-CEX has been adopted for a variety of assessment purposes and is now not only used in the US but also in other countries such as Canada, the United Kingdom, Australia and Argentina. It has been suggested that the Mini-CEX may be the "only evaluation method used by many residency programs to directly observe clinical skills" (Holmboe et al. 2003, p. 826). The Mini-CEX is also used to assess residents in other specialties and its use has extended to other examinee groups such as undergraduate medical students (Dewi and Achmad 2010; Hill and Kendall 2007; Hill et al. 2009; Kogan et al. 2003; Lie et al. 2010; Ney et al. 2009), practicing doctors (Sidhu et al. 2009), and international medical graduates (Nair et al. 2008). In addition to being recommended by ABIM and ACGME, its use is also recommended by other regulators and governing bodies. As examples, the Postgraduate Medical Education and Training Board in the United Kingdom recommends the use of the Mini-CEX for assessment in the postgraduate setting (Hill et al. 2009), the Mini-CEX is mandatory during dermatology specialist training in the UK (Cohen et al. 2009), and the Australian Medical Council has introduced the Mini-CEX as a workplace assessment tool for some international medical graduates (Nair et al. 2008). In addition to providing formative feedback to guide further education and training, the Mini-CEX has been used for summative purposes to make educational decisions about medical students (Hill et al. 2009) and residents (Weller et al. 2009).

Systematic Reviews of the Mini-CEX

Two studies have used systematic reviews to investigate validity evidence for direct observation assessment methods including the Mini-CEX. Kogan et al. (2009) identified 55 tools used for direct observation and assessment and investigated evidence of their validity and outcomes. They concluded that the Mini-CEX is one of few tools that has been thoroughly evaluated and that it has the strongest validity evidence of the 55 assessment tools they investigated. However Pelgrim et al. (2010) also studied multiple direct observation tools and concluded that although the validity of the Mini-CEX is supported by correlations with other assessment instruments, additional types of validity evidence are lacking.

A third systematic review conducted by Hawkins et al. (2010) focussed specifically on the Mini-CEX and analyzed validity evidence within the framework of a validity argument (Kane 1992). Hawkins et al. (2010) found that there are relatively few studies of the Mini-CEX, the studies that do exist have variable designs that

present conflicting results, and it is "difficult to separate problems with the method from gaps and limitations in the research conducted to date." (p. 1495)

These three systematic reviews present conflicting views of the state of Mini-CEX validity research and evidence. Taken together, they raise questions about the degree and types of validity evidence that may support use of Mini-CEX scores and they highlight the need to examine potential gaps and limitations in the Mini-CEX validation research. As we noted, one way of doing this is to examine the degree to which the body of Mini-CEX validation research is aligned with contemporary validity theory.

Contemporary Validity Theory

The *Standards* (AERA et al. 1999) define validity as follows:

> Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each intended interpretation must be validated.
>
> Validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use. The proposed interpretation refers to the construct or concepts the test is intended to measure. (p. 9)

Two aspects of the current view of validity require particular emphasis for the purposes of this paper. First, evidence for validity ought to be of highest priority to test developers, users and researchers because validity is *the most fundamental consideration* in developing and evaluating tests. In the contemporary view of validity other evidence such as reliability evidence contributes to validity and is necessary but insufficient for defensible use of test scores. Therefore validity evidence is required in addition to evidence for other characteristics such as reliability, feasibility or utility (the latter two being often reported in medical education literature).

Second, since validity pertains to interpretations and uses of test scores and not to tests themselves, validation efforts should be focussed on the proposed interpretations and uses of test scores and should begin by clearly specifying what the interpretations and uses are. In a contemporary view validation efforts consist of two inter-related arguments: an interpretive argument and a validity argument (Cronbach 1988; Hubley and Zumbo 1996; Kane 1992, 2001, 2013; Messick 1989). The interpretative argument specifies the proposed interpretations and intended uses of the test or assessment scores by identifying inferences and assumptions that flow from them, while the validity argument systematically evaluates the interpretive argument. When a particular assessment or test is used

in more than one setting for more than one application, the inferences and assumptions may change and the evidence required to support them may change. Nonetheless, validation will involve the specification (the interpretive argument) and evaluation (the validity argument) of the proposed interpretations and uses of the scores. Thus, claims about validity of test interpretations and uses are claims that the interpretive argument, the inferences and the assumptions are logical and plausible in the application in which the scores are being used (Kane 2006, 2013).

As set out in the *Standards*, specifying the interpretation begins with adequately defining the construct being measured. A construct is a broad term for the concept or characteristic a test is designed to measure, and the purpose of a test is to make inferences from test scores to unobservable constructs such as knowledge, ability, aptitude or competence. All tests should be construct-referenced because the interpretation of the construct is the foundation for the score-based inferences that arise from test use (Messick 1989). Test use and validation must proceed by clearly and thoroughly defining the construct being measured. Simply naming or labelling the construct is insufficient because the same name or label can be applied to different constructs – a common name does not automatically imply a common construct (Reckase 1998). As an example, the construct of "clinical competence" takes on different meanings when used by different parties or in different settings. Attempts to validate assessments of clinical competence should begin with a clear understanding of what is meant by clinical competence in the setting in which the assessment instrument will be used. Once the construct and proposed interpretation and inferences have been identified, evaluation through the use of a validity argument proceeds by developing empirical evidence, examining relevant literature, and/or conducting logical analyses.

Sources of Validity Evidence

The contemporary view of validity and validation requires validity evidence to be integrated from multiple sources to develop the validity argument that supports intended uses and interpretations of scores and to rule out threats to validity (Messick 1989, 1994). The *Standards* outline five sources of validity evidence that should be investigated for these purposes.

Evidence Based on Test Content

Evidence for validity can be found by analyzing the relationship between the test content and the construct intended to be measured. Sireci (1998) noted that content validity involves four commonly-accepted elements: domain definition (the conceptual and operational definitions of the construct); domain representation (match between a test and the domain definition); domain relevance (relevance of items to the content domain); and appropriate test construction procedures. Evidence based

on test content can be sought through logical or empirical analyses, including the use of subject matter experts to examine the theoretical relationship between the construct and the test content, write test items, and review item specifications, test blueprints and documentation.

Evidence Based on Response Processes

"Response processes" refers to the detailed characteristics of performance or response actually engaged in by examinees or examiners during the assessment event. Evidence based on response processes provides information about the fit between the construct and the cognitive processes engaged in during a test. For example, in a test of clinical reasoning, evidence would be required to determine whether examinees are actually using clinical reasoning skills (as opposed to perhaps following a memorized pattern of response). Evidence based on response processes can be gathered by questioning test-takers or examiners about their strategies or responses through the use of surveys, interviews, or think-aloud procedures and expert review (Miller and Linn 2000).

Evidence Based on Internal Structure

Internal structure refers to relationships between items or parts of a test. Information about a test's internal structure can reveal how closely the test conforms to the construct of interest. For example, if a test is intended to measure a unidimensional construct, then evidence of structural unidimensionality would support the relationship between the test and the construct, or if the construct is thought to be composed of several components, then multidimensionality in the test's internal structure would support that. Methods of gathering evidence based on internal structure include examining the factor structure of the data through confirmatory factor analysis, and conducting differential item functioning analyses to determine whether test items may behave differently for subgroups of examinees.

Evidence Based on Relations to Other Variables

Evidence based on relationships with other variables provides information about the extent to which the relationships are consistent with the intended construct. *Convergent validity evidence* is gathered by examining relationships between the test scores and other measures that are intended to assess theoretically-similar constructs, whereas *discriminant validity evidence* is drawn by examining relationships with measures intended to assess theoretically-different constructs. According to the *Standards*, group membership variables are relevant if the theory underlying the test use suggests that group differences should be present. For example, studies that show that scores are higher for more experienced examinees than for less

experienced examinees (or for instructed versus non-instructed examinees) provide convergent validity evidence because there is a theoretical basis for expecting score differences between the groups. *Test-criterion validity evidence* examines how accurately test scores predict a criterion performance where the criterion variable is an attribute or outcome of interest. A concurrent test-criterion study collects data from the predictor and criterion measures at approximately the same time, whereas in a predictive test-criterion study the criterion scores are obtained after the predictor scores. *Validity generalization* evidence refers to the degree to which evidence of validity based on test-criterion relations can be generalized to a new situation, for example through the use of meta-analysis. Evidence based on relations to other variables can be assessed through experimental and correlational studies, or through a multitrait-multimethod matrix approach (Campbell and Fiske 1959).

Evidence Based on Consequences of Testing

Although there is debate on this topic, evidence about the intended and unintended consequences of test use is currently required by the *Standards*. Therefore it is important to investigate whether intended consequences are occurring as anticipated, or whether *un*intended consequences may be occurring. For example, when a claim is made that a formative assessment has a positive impact on learning (such as the case of the Mini-CEX where a critical component of the assessment is the provision of feedback to examinees for the purpose of improving their performance), the validation process should question whether the positive impact is being realized.

There has been some deliberation in the literature as to whether all types of validity evidence are required for all types of assessments. The current position expressed in the *Standards* is that some sources of evidence will be especially important to evaluate in a given case, yet strong evidence from one source does not diminish the need for evidence from other sources. Therefore evidence from all five sources should be found within the body of Mini-CEX research, although they may be found to varying degrees.

Method

We conducted a search for English language literature published between January 1995 (the year in which the Mini-CEX was first introduced) and December 31, 2012 in Academic Search Complete, CINAHL, Education Research Complete, ERIC, MEDLINE, and PsychINFO. The search terms used were "Mini-Clinical Evaluation Exercise" or "Mini-CEX" and "valid*" (to capture valid, validity and validation) in all text. From this initial search we removed duplicates and excluded publications if they: (1) were not primary research, or were summaries, reviews,

interpretations or critiques of prior research; (2) did not investigate aspects of the Mini-CEX (for example, articles whose main purpose was to investigate other assessment tools but also mentioned the Mini-CEX); or (3) were editorials, letters to the editor, or conference abstracts. In addition, we examined the references in review articles to ensure the search was as comprehensive as possible.

To determine whether the main intent of each study was to present validity evidence (i.e., is the study a validity study of the Mini-CEX?), we coded whether any of the words "valid", "validity" or "validation" appeared in the title, abstract or key words and descriptors pertaining to the study. If they did we coded the study as a "validity study" and if not we coded the study as a "non-validity study".

To address the first research question regarding the extent to which the validity studies present views of validity that align with the contemporary view of validity theory, we coded whether each validity study: (1) presented a definition of validity similar to the *Standards*; (2) made reference to either the *Standards* or to contemporary validity theorists (such as those that would be taught in an introductory validity course); (3) identified and defined the construct being assessed; (4) presented a view of validity as a characteristic of Mini-CEX scores and inferences rather than as a characteristic of the Mini-CEX; (5) described the use of the Mini-CEX (for example, described the population being assessed in terms of their level of education and specialty where appropriate, the setting in which the assessment was taking place and whether the Mini-CEX scores were intended to provide formative or summative assessment information); and (6) described the intended interpretation and inferences to be drawn from Mini-CEX assessment results.

To address the second research question about the extent to which the recommended sources of validity evidence outlined in the *Standards* has been reported regarding the Mini-CEX, we coded the sources or types of validity evidence reported in the validity studies. To allow a comparison between the validity perspective taken in the studies and the validity perspective of the *Standards* and to investigate whether the sources of validity evidence being reported were aligned with sources of validity evidence in the *Standards* we re-coded the type of evidence reported in the studies as it would be reported according to the *Standards* framework. In addition, if validity evidence was presented in the non-validity studies we coded it also according to the *Standards* framework. This allowed us to fully address our second research question and determine the extent to which all recommended sources of validity evidence have been reported in all published studies of the Mini-CEX regardless of whether the studies were presented as validity studies of the Mini-CEX or not.

For both validity and non-validity studies we coded other measurement characteristics that were reported such as reliability, feasibility, utility and acceptability. Further, we coded the types of reliability evidence reported (including alternate forms, test-retest, internal consistency, scorer consistency, G-theory reproducibility, standard errors of measurement, or item response theory test information function).

All coding was carried out by the first author. In order to investigate accuracy of the coding procedure we calculated inter-rater reliability. Another researcher familiar with medical education research and contemporary validity theory coded 6 - randomly-selected studies. First, we explained the purpose of this study and reviewed the coding sheet with her. She then coded the studies independently and without knowledge of the first author's coding results.

Results

After excluding articles that did not meet the inclusion criteria as set out above, 43 articles were included in this study. A list of the included studies is attached as Appendix. Thirteen of the 43 included studies appeared to be positioned as Mini-CEX validity studies and they comprise the validity studies group. That is, 13 studies investigated the properties of the Mini-CEX and used the word "valid", "validity", or "validation" in the title, abstract or key words/descriptors pertaining to the study. The remaining 30 studies comprise the non-validity studies group.

Figure 17.1 shows the distribution of all included studies according to the year they were published. The first validity study of the Mini-CEX was published in 2002, 7 years after its inception. The majority of validity and non-validity studies have been published since 2006.

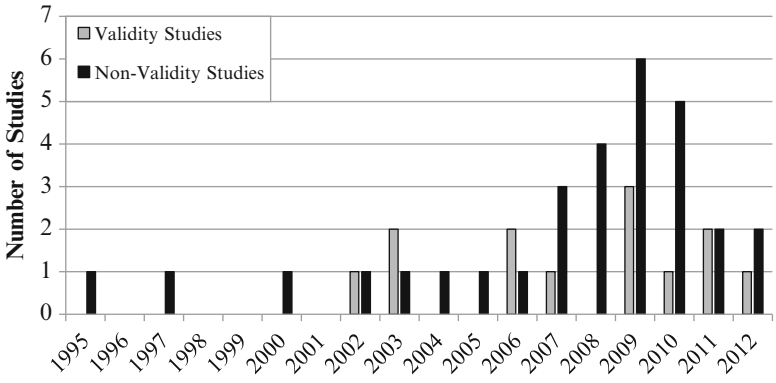

Fig. 17.1 Number of Mini-CEX studies published per year

Validity Studies' Alignment with Contemporary Validity Theory

The results of coding the 13 validity studies to determine their alignment with contemporary validity theory are summarized as follows.

None of the validity studies presented a definition of validity similar to the *Standards* although one defined "construct validity". None of the validity studies made reference to the *Standards* or to validity theorists directly, although one validity study cited an article that summarizes the *Standards* and the contemporary view of validity theory. Two validity studies provided limited (one or two sentences) definitions of the construct intended to be assessed and one study provided a reference to documentation where the construct was defined. Ten of the validity studies did not define the construct being assessed. Twelve of the 13 validity studies named a construct: 3 were reported as "competence", 4 as "clinical skills" and 5 as "clinical competence". Most validity studies named the skills that were assessed (such as history taking or physical examination) however none referred to any theoretical relationship between the skills assessed and the construct. Five validity studies clearly characterized validity as a property of the test, 5 as a property of scores or inferences, and 5 were unclear.

Figure 17.2 shows the uses of the Mini-CEX reported in the validity studies broken down by educational level, medical specialty, and assessment type. This figure reveals that for the most part the settings in which the Mini-CEX has been studied have been reported in the validity studies. As can be expected from the history of the Mini-CEX, most validity studies have investigated its use in internal medicine residencies as a form of formative assessment, although validity evidence has also been gathered for other uses and in other settings. Please note that some studies reported more than one use therefore the totals add up to more than the number of studies.

We also coded whether each validity study described how the Mini-CEX scores were to be interpreted and the inferences to be drawn from them in the particular setting of the study. No validity study specifically described the interpretation and

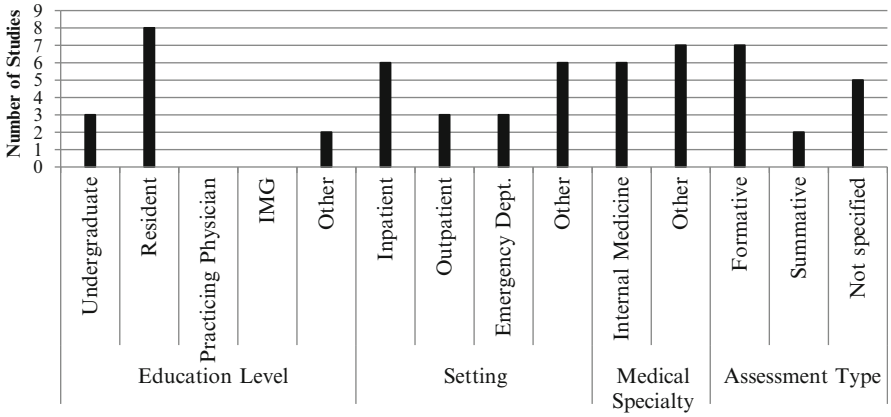

Fig. 17.2 Uses of the Mini-CEX reported in the validity studies

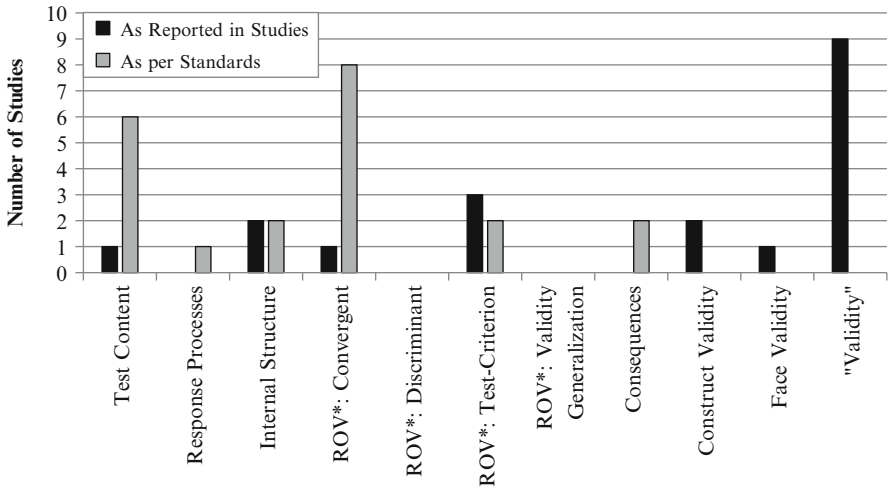

Fig. 17.3 Sources of validity evidence in the validity studies (**ROV* Relations to other Variables)

inferences that were to be drawn from the Mini-CEX scores in the particular setting of the study. Although a few studies touched on the issue, in most studies it was implicit that simply stating whether the assessment was formative or summative was sufficient to deduce whatever inferences were to be drawn.

Sources of Validity Evidence Reported in the Validity Studies

Figure 17.3 shows the sources of validity evidence as reported in the validity studies, and contrasts how the sources of validity evidence were presented in the validity studies with how the same evidence would be framed within the *Standards* framework.

Three of the 13 validity studies presented validity evidence similarly to the *Standards*; however, as can be seen in Fig. 17.3, there are considerable differences between study perspectives and *Standards'* perspectives as to sources of validity evidence in the remaining studies. Of the 9 studies that presented unspecified sources of validity evidence (i.e. evidence was referred to simply as "validity"), 4 presented evidence based on convergent relations to other variables, 4 presented test content validity evidence, 1 presented response process validity evidence, 1 presented test criterion evidence, and 2 presented evidence related to consequences of testing. In addition, 2 studies that presented construct validity evidence and 2 that presented criterion evidence were recoded as presenting evidence based on convergent relations to other variables. None of the validity studies presented evidence related to discriminant relations with other variables or validity generalization. Please note that the total number of sources of validity evidence presented is greater than the total number of validity studies because some studies presented more than one type of validity evidence.

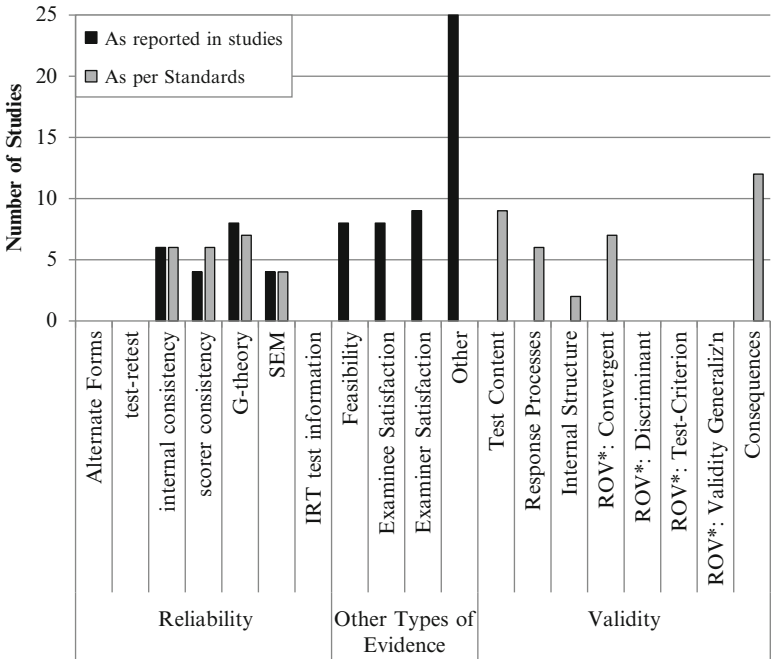

Fig. 17.4 Types of evidence presented in the non-validity studies (**ROV* Relations to other Variables)

Other Types of Evidence Reported in the Validity Studies

Of the validity studies that presented reliability evidence, five presented internal consistency evidence, three presented scorer consistency evidence, five presented generalizability theory reproducibility evidence, and two presented standard error of measurement (SEM). In addition, five of the validity studies presented feasibility evidence, four presented evidence of examinee satisfaction and four presented evidence of examiner satisfaction.

Types of Evidence Presented in the Non-validity Studies

Types of evidence presented in the 30 non-validity studies are shown in Fig. 17.4 which also contrasts how evidence was presented in the non-validity studies with how the same evidence would be framed according to the *Standards*. Most (22) of the non-validity studies reported reliability evidence and many also reported feasibility, examinee satisfaction, and examiner satisfaction. Twenty-five of the non-validity studies reported a variety of other properties of the Mini-CEX.

Examples of terms used to describe other properties were utility, accuracy, psychometric characteristics, use, acceptability, and influence on feedback. Since most studies reported more than one type of evidence the total number is greater than 30.

As in the case of the validity studies, the evidence presented for the non-validity studies would be classified differently when viewed from the perspective of the *Standards*. For the most part, reliability evidence has been framed in the studies similarly to the way it would be framed according to the *Standards* as evidenced by the similar patterns for reliability in Fig. 17.4. However, also shown in Fig. 17.4, a considerable amount of validity evidence was presented in the non-validity studies yet was not identified in the studies as validity evidence. For example, one study investigated the Mini-CEX in terms of its educational impact, the factors that influence examiner scoring decisions, and its effects on the relationship between examiner and examinee (amongst other things). According to the *Standards* these types of investigations provide information about validity such as evidence related to response processes and consequences of testing. The main types of validity evidence presented in the non-validity studies were validity evidence based on test content, response processes, convergent relations to other variables, and consequences of testing.

To What Extent Have the Recommended Sources of Validity Evidence Outlined in the Standards *Been Reported Regarding the Mini-CEX?*

Figure 17.5 shows all sources of validity evidence stemming from the 43 validity and non-validity studies combined, categorized as per the *Standards*. This figure reveals that the combined Mini-CEX research efforts (when conceptualized aligned with contemporary validity theory) have focussed predominantly on validity evidence based on test content, response processes, convergent relations to other variables, and consequences of testing. To date, the body of Mini-CEX validation research does not provide evidence based on discriminant relations to other variables or validity generalization.

Inter-rater Agreement on Coding of the Studies

Six randomly selected studies were rated by an independent rater to investigate accuracy of the coding procedure used by the first author of the study. The independent rater and first author were in agreement on 381 of the 438 total data points on the coding sheets for the 6 studies, representing 87 % inter-rater agreement. Differences were discussed and reviewed until agreement was reached.

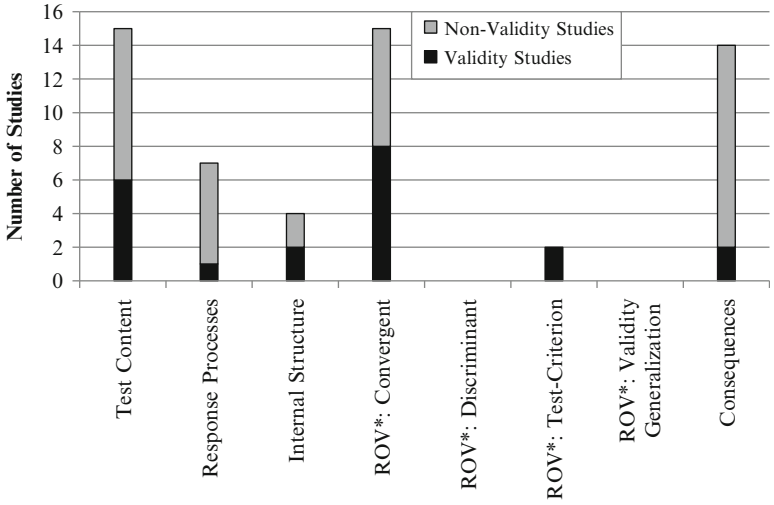

Fig. 17.5 Sources of validity evidence investigated in the validity and non-validity studies (**ROV* Relations to other Variables)

Discussion and Conclusions

This study is the first study to examine the extent to which research that investigates validity evidence for the Mini-CEX conforms to contemporary validity theory and meets the recommendations and criteria set out in the *Standards for Educational and Psychological Testing* (AERA et al. 1999). It is not the intent of this research to assess or comment on the quality of the research reviewed, rather to understand and report on the validity perspective taken in the body of literature regarding Mini-CEX and to determine which sources of validity evidence have been investigated. The results provide interesting findings about the manner in which validity is conceptualized and presented in the Mini-CEX literature and point to gaps and limitations in the research.

This study provides evidence that the body of validity research of the Mini-CEX is not fully aligned with contemporary validity theory because it does not place emphasis on the proposed interpretations and uses of test scores and on the theoretical relationship between the test and the construct being assessed. As stated in the *Standards*, validation efforts should begin with a clear definition of the proposed interpretation of scores which refers to the construct intended to be measured, together with a rationale that connects the interpretation to the proposed use of the scores. Results of the current study indicate that these first steps in the validation process for the Mini-CEX have not yet been taken: most of the validity studies investigated in this research do not provide a definition of the construct being assessed or a theoretical rationale to guide the interpretation of Mini-CEX scores.

Although most of the validity studies adequately provide contextual information about their local use of the Mini-CEX (such as the education level of the examinees, the setting in which the Mini-CEX was administered and the medical specialty that was being assessed), there was very little information about how the scores were intended to be interpreted and used. Although a few studies touched on the issue, in most studies it was implicit that simply stating whether the assessment was formative or summative was sufficient to deduce the inferences to be drawn. For example, one study reported that the Mini-CEX evaluations did not contribute to final grades of the examinees but the actual interpretations and uses of the assessments were not stated. The findings of this study confirm those of Hawkins et al. (2010) who noted that a lack of attention in the Mini-CEX validation literature to Mini-CEX use and score interpretations is a concern.

Of the studies that were presented as validity studies, none provided a definition of validity and many framed validity evidence differently than it would have been if it were aligned with the *Standards*. Only five of the validity studies characterized validity as a property of test scores or inferences. The remaining validity studies were either unclear in their position or explicitly referred to validity as a property of the Mini-CEX. For example, phrases such as "the Mini-CEX has construct validity" or "the validity of the Mini-CEX" were frequently observed in the studies. As early as the 1974 edition of the *Standards* it was considered incorrect to use the unqualified phrase "the validity of the test" (Sireci 2009) yet the results of this study point to evidence that this terminology and characterization of validity still exists in the body of research about the Mini-CEX.

This study also provides evidence about which sources of validity evidence have been reported in the Mini-CEX literature. Most studies to date have focussed on evidence based on convergent relations to other variables, test content, and consequences of testing. Few have focussed on response processes, internal structure, and test-criterion, and no studies have investigated validity evidence based on discriminant relations to other variables or validity generalization. Much of the validity evidence has arisen from studies that were not presented as having validity as their major focus and some validity evidence has been presented using other terminology such as feasibility, utility or acceptability with no connection being made to validity or validity theory. These findings support those of Pelgrim et al. (2010) who reported that few sources of validity evidence have been addressed in Mini-CEX research. They also support the findings of Hawkins et al. (2010) who found gaps and limitations in Mini-CEX validation research.

One way in which the Mini-CEX validation research is aligned with contemporary validity theory is that it is an ongoing endeavour with much research activity over the last 5 years. This practice is aligned with the *Standards* which set out that validation is a continual process and that as new uses of an assessment tool arise (as they have in the case of the Mini-CEX), research should continue to investigate sources of validity evidence associated with new use.

Implications of Findings and Suggestions for Future Research

The findings that sources of validity evidence are conceptualized differently in the published Mini-CEX validation literature than in the *Standards* and that many of the published studies presented validity evidence outside of a validity framework have implications for researchers and for journal editors. Future research could focus on enhancing researcher awareness and rectifying misunderstandings about what to report as validity evidence and how to report it. Journal editors may consider setting clear inclusion and exclusion guidelines and strengthening the peer review process for studies that investigate psychometric properties of assessments such as the Mini-CEX. Further, we found that not all studies that present validity evidence have any form of the word "valid" in their title, abstract, key words or descriptors thus making it difficult for future researchers to find the validation research that does exist. Researchers and journal editors may address this shortcoming to ensure that all future validation research will be readily accessible through typical search strategies and thus play an important role in disseminating key validity information.

The results of this study also have implications for Mini-CEX users such as medical education programs and governing bodies that set policy that recommends or mandates its use. They should be aware of the gaps in the research and degree of alignment or lack of alignment with the *Standards* and carefully consider the extent to which the existing literature supports their recommendations or the inferences to be drawn from their particular use of Mini-CEX thus ensuring that their recommendations and uses are defensible. As noted in the introduction, validation research that is closely aligned with the *Standards* most strongly supports defensible use and interpretations of test scores (Sireci and Parker 2006).

Perhaps the most important implications from this research derive from the finding that to date Mini-CEX validation research neither provides a *theoretical* rationale for score interpretation and use based on a clearly-understood construct nor clearly elucidates the inferences to be drawn from Mini-CEX use. This finding leads us to conclude that the body of Mini-CEX validation research as a whole currently represents a "weak program" of validation research (Cronbach 1988), that is, one that presents validity evidence without reference to theoretical underpinnings and often relies on data that is easily or readily available as opposed to data that is relevant (Kane 2001). Further, as noted by Kane, as early as the 1970s there was concern about the ease with which opportunistic validity evidence could be presented without stating a proposed interpretation or evaluating the reasonableness of the interpretation. In other words, the two key elements of the validation process (a clearly-stated interpretive argument and a validity argument which evaluates it) are deficient in a weak program of validation. A strong theory-driven program of research which will assure scientific and disciplined enquiry (Zumbo 2009) requires multiple strands of evidence some based on statistical analyses and some based on theory (Sireci 2009).

When a field or area of research is inhibited by the absence of well-defined theory about the construct a strong program of validity research will be difficult to achieve. It is important to note that work is being done which will help to develop theory about the construct being assessed by the Mini-CEX. For example, our study revealed research that provided validity evidence based on response processes which is theory-building research (see, as examples, Kogan et al. 2011, 2012; Weller et al. 2009). However, such research is being conducted outside of a contemporary validity theory framework. Indeed, our study revealed that a great deal of validity evidence related to the Mini-CEX has been presented in studies that fail to make any connection whatsoever to validity. A lack of connection from the research to validity or validity theory weakens or undermines the ability to develop a sound validity argument.

Kane (2001) draws distinctions between performance assessments of observable attributes and those of theoretical constructs and notes that clearly defined observable attributes might be validated with relatively simple interpretive arguments and clear validation strategies without reference to underlying theories about what is being assessed. However, the extent to which the intended interpretations generalize or go beyond the observations being made determines the strength of validity argument required: in the case of the Mini-CEX, if the intended interpretation extends from observed scores to more general conclusions about competence, then a strong program of validity research should be required. If not, a weaker program based on readily-available data may suffice. Regardless of whether the Mini-CEX is construed as assessing a theoretical construct or an observable attribute, future validation research may be directed at defining what is being assessed, building the *theoretical* rationale for score interpretation and clarifying the inferences to be drawn from Mini-CEX use thereby contributing to a stronger body of Mini-CEX validation research than currently exists.

Appendix: List of Included Studies

Alves de Lima, A., Henquin, R., Thierer, J., Paulin, J., Lamari, S., Belcastro, F., & Van der Vleuten, C. P. M. (2005). A qualitative study of the impact on learning of the mini clinical evaluation exercise in postgraduate training. *Medical Teacher*, *27* (1), 46–52.

Alves de Lima, A., Barrero, C., Baratta, S., Castillo Costa, Y., Bortman, G., Carabajales, J., Conde, D., Galli, A., Degrange, G., & van der Vleuten, C. (2007). Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Medical Teacher*, *29*(8), 785–790.

Alves de Lima, A. E., Conde, D., Aldunate, L., & van der Vleuten, C. P. (2010). Teachers' experiences of the role and function of the mini clinical evaluation exercise in post-graduate training. *International Journal of Medical Education*, *1*, 68–73.

Brazil, V., Ratcliffe, L., Zhang, J., & Davin, L. (2012). Mini-CEX as a workplace-based assessment tool for interns in an emergency department – Does cost outweigh value? *Medical Teacher*, *34*(12), 1017–1023. doi:10.3109/0142159X.2012.719653.

Chen, W., Lai, M.-M., Li, T.-C., Chen, P. J., Chan, C.-Y., & Lin, C.-C. (2011). Professional development is enhanced by serving as a mini-CEX preceptor. *Journal of Continuing Education in the Health Professions*, *31*(4), 225–230.

Cohen, S. N., Farrant, P. B. J., & Taibjee, S. M. (2009). Assessing the assessments: UK dermatology trainees' views of the workplace assessment tools. *British Journal of Dermatology*, *161*(1), 34–39.

Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education*, *14*(5), 655–664.

Cook, D. A., Dupras, D. M., Beckman, T. J., & Thomas, K. G. (2009). Effect of rater training on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *Journal of General Internal Medicine*, *24*(1), 74–79.

Cook, D. A., Beckman, T. J., Mandrekar, J. N., & Pankratz, V. S. (2010). Internal structure of mini-CEX scores for internal medicine residents: Factor analysis and generalizability. *Advances in Health Sciences Education*, *15*(5), 633–645.

Crossley, J., Johnson, G., Booth, J., & Wade, W. (2011). Good questions, good answers: Construct alignment improves the performance of workplace-based assessment scales. *Medical Education*, *45*(6), 560–569.

Dewi, S. P., & Achmad, T. H. (2010). Optimising feedback using the mini-CEX during the final semester programme. *Medical Education*, *44*(5), 509–509.

Durning, S. J., Cation, L. J., Markert, R. J., & Pangaro, L. N. (2002). Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. *Academic Medicine*, *77*(9), 900.

Fernando, N., Cleland, J., McKenzie, H., & Cassar, K. (2008). Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. *Medical Education*, *42*(1), 89–95.

Hatala, R., Ainslie, M., Kassen, B. O., Mackie, I., & Roberts, J. M. (2006). Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. *Medical Education*, *40*(10), 950–956.

Hauer, K. E. (2000). Enhancing feedback to students using the Mini-CEX (Clinical Evaluation Exercise). *Academic Medicine*, *75*(5), 524.

Hill, F., & Kendall, K. (2007). Adopting and adapting the mini-CEX as an undergraduate assessment and learning tool. *The Clinical Teacher*, *4*(4), 244–248.

Hill, F., Kendall, K., Galbraith, K., & Crossley, J. (2009). Implementing the undergraduate mini-CEX: A tailored approach at Southampton University. *Medical Education*, *43*(4), 326–334.

Holmboe, E. S., Huot, S., Chung, J., Norcini, J., & Hawkins, R. E. (2003). Construct validity of the miniClinical Evaluation Exercise (miniCEX). *Academic Medicine*, *78*(8), 826.

Holmboe, E. S., Yepes, M., Williams, F., & Huot, S. J. (2004). Feedback and the Mini Clinical Evaluation Exercise. *Journal of General Internal Medicine*, *19*(5p2), 558–561.

Jackson, D., & Wall, D. (2010). An evaluation of the use of the mini-CEX in the foundation programme. *British Journal of Hospital Medicine*, *71*(10), 584–588.

Kogan, J. R., & Hauer, K. E. (2006). Brief report: Use of the Mini-Clinical Evaluation Exercise in internal medicine core clerkships. *Journal of General Internal Medicine*, *21*(5), 501–502.

Kogan, J. R., Bellini, L. M., & Shea, J. A. (2002). Implementation of the mini-CEX to evaluate medical students' clinical skills. *Academic Medicine*, *77*(11), 1156–1157.

Kogan, J. R., Bellini, L. M., & Shea, J. A. (2003). Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Academic Medicine*, *78*(10), S33–S35.

Kogan, J. R., Conforti, L., Bernabeo, E., Iobst, W., & Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education*, *45*(10), 1048–1060.

Kogan, J. R., Conforti, L. N., Bernabeo, E. C., Durning, S. J., Hauer, K. E., & Holmboe, E. S. (2012). Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*, *46*(2), 201–215.

Lie, D., Encinas, J., Stephens, F., & Prislin, M. (2010). Do faculty show the "halo effect" in rating students compared with standardized patients during a clinical examination. *The Internet Journal of Family Practice*, *8*(2). Retrieved from http://www.ispub.com/journal/the_internet_journal_of_family_practice/volume_8_number_2_20/article/do-faculty-show-the-halo-effect-in-rating-students-compared-with-standardized-patients-during-a-clinical-examination.html

Malhotra, S., Hatala, R., & Courneya, C.-A. (2008). Internal medicine residents' perceptions of the Mini-Clinical Evaluation Exercise. *Medical Teacher*, *30*(4), 414–419.

Margolis, M. J., Clauser, B. E., Cuddy, M. M., Ciccone, A., Mee, J., Harik, P., & Hawkins, R. E. (2006). Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: A validity study. *Academic Medicine*, *81*(10), S56–S60.

Mitchell, C., Bhat, S., Herbert, A., & Baker, P. (2011). Workplace-based assessments of junior doctors: Do scores predict training difficulties? *Medical Education*, *45*(12), 1190–1198.

Nair, B. R., Alexander, H. G., McGrath, B. P., Parvathy, M. S., Kilsby, E. C., Wenzel, J., Frank, I. B., Pachev, G. S., & Page, G. G. (2008). The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *The Medical Journal of Australia*, *189*(3), 159–161.

Ney, E. M., Shea, J. A., & Kogan, J. R. (2009). Predictive validity of the mini-Clinical Evaluation Exercise (mCEX): Do medical students' mCEX ratings correlate with future clinical exam performance? *Academic Medicine*, *84*(10), S17–S20.

Norcini, J. J., Blank, L. L., Arnold, G. K., & Kimball, H. R. (1995). The Mini-CEX (Clinical Evaluation Exercise): A preliminary investigation. *Annals of Internal Medicine*, *123*(10), 795–799.

Norcini, J. J., Blank, L. L., Arnold, G. K., & Kimball, H. R. (1997). Examiner differences in the mini-CEX. *Advances in Health Sciences Education*, *2*(1), 27–33.

Norcini, J. J., Blank, L. L., Duffy, F. D., & Fortna, G. S. (2003). The mini-CEX: A method for assessing clinical skills. *Annals of Internal Medicine*, *138*(6), 476–481.

Ogunbanjo, G. A. (2009). Adapting mini-CEX scoring to improve inter-rater reliability. *Medical Education*, *43*(5), 484–485.

Quantrill, S. J., & Tun, J. K. (2012). Workplace-based assessment as an educational tool. Guide supplement 31.5-Viewpoint. *Medical Teacher*, *34*(5), 417–418.

Sidhu, R. S., Hatala, R., Barron, S., Broudo, M., Pachev, G., & Page, G. (2009). Reliability and acceptance of the Mini-Clinical Evaluation Exercise as a performance assessment of practicing physicians. *Academic Medicine*, *84*(10), S113–S115.

Torre, D. M., Simpson, D. E., Elnicki, D. M., Sebastian, J. L., & Holmboe, E. S. (2007). Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. *Teaching and Learning in Medicine*, *19*(3), 271–277.

Van Lohuizen, M. T., Kuks, J. B., van Hell, E. A., Raat, A. N., Stewart, R. E., & Cohen-Schotanus, J. (2010). The reliability of in-training assessment when performance improvement is taken into account. *Advances in Health Sciences Education*, *15*(5), 659–669.

Weller, J. M., Jolly, B., Misur, M. P., Merry, A. F., Jones, A., Crossley, J. M., Pedersen, K., & Smith, K. (2009a). Mini-clinical evaluation exercise in anaesthesia training. *British Journal of Anaesthesia*, *102*(5), 633–641.

Weller, J. M., Jones, A., Merry, A. F., Jolly, B., & Saunders, D. (2009b). Investigation of trainee and specialist reactions to the mini-Clinical Evaluation Exercise in anaesthesia: Implications for implementation. *British Journal of Anaesthesia*, *103*(4), 524–530.

Wiles, C. M., Dawson, K., Hughes, T. A. T., Llewelyn, J. G., Morris, H. R., Pickersgill, T. P., Robertson, N. P., & Smith, P. E. M. (2007). Clinical skills evaluation of trainees in a neurology department. *Clinical Medicine*, *7*(4), 365–369.

Wilkinson, J. R., Crossley, J. G., Wragg, A., Mills, P., Cowan, G., & Wade, W. (2008). Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical Education*, *42*(4), 364–373.

References

American Board of Internal Medicine. (2009). *Assessment tools*. Retrieved November 20, 2010, from http://www.abim.org/program-directors-administrators/assessment-tools/mini-cex.aspx

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*(3), 397–412.

Cizek, G. J., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement, 70*(5), 732–743.

Cohen, S. N., Farrant, P. B. J., & Taibjee, S. M. (2009). Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *British Journal of Dermatology, 161*(1), 34–39.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillside: Lawrence Erlbaum.

Dewi, S. P., & Achmad, T. H. (2010). Optimising feedback using the mini-CEX during the final semester programme. *Medical Education, 44*(5), 509–509.

Hawkins, R. E., Margolis, M. J., Durning, S. J., & Norcini, J. J. (2010). Constructing a validity argument for the mini-clinical evaluation exercise: A review of the research. *Academic Medicine: Journal of the Association of American Medical Colleges, 85*(9), 1453–1461.

Hill, F., & Kendall, K. (2007). Adopting and adapting the mini-CEX as an undergraduate assessment and learning tool. *The Clinical Teacher, 4*(4), 244–248.

Hill, F., Kendall, K., Galbraith, K., & Crossley, J. (2009). Implementing the undergraduate mini-CEX: A tailored approach at Southampton University. *Medical Education, 43*(4), 326–334.

Hogan, T. P., & Agnello, J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational and Psychological Measurement, 64*(5), 802–812.

Holmboe, E. S., Huot, S., Chung, J., Norcini, J., & Hawkins, R. E. (2003). Construct validity of the MiniClinical evaluation exercise (MiniCEX). *Academic Medicine, 78*(8), 826–830.

Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology, 123*, 207–215.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527–535.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319–342.

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.). Washington, DC: American Council on Education.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73. doi:10.1111/jedm.12000.

Kogan, J. R., Bellini, L. M., & Shea, J. A. (2003). Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Academic Medicine: Journal of the Association of American Medical Colleges, 78*(10), S33–S35.

Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. *JAMA, the Journal of the American Medical Association, 302*(12), 1316–1326.

Kogan, J. R., Conforti, L., Bernabeo, E., Iobst, W., & Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education, 45*(10), 1048–1060.

Kogan, J. R., Conforti, L. N., Bernabeo, E. C., Durning, S. J., Hauer, K. E., & Holmboe, E. S. (2012). Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education, 46*(2), 201–215.

Lie, D., Encinas, J., Stephens, F., & Prislin, M. (2010). Do faculty show the 'halo effect' in rating students compared with standardized patients during a clinical examination? *Internet Journal of Family Practice, 8*(2), 1–1.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Miller, M. D., & Linn, R. L. (2000). Validation of performance-based assessments. *Applied Psychological Measurement, 24*(4), 367.

Nair, B. R., Alexander, H. G., McGrath, B. P., Parvathy, M. S., Kilsby, E. C., Wenzel, J., et al. (2008). The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *The Medical Journal of Australia, 189*(3), 159–161.

Ney, E. M., Shea, J. A., & Kogan, J. R. (2009). Predictive validity of the mini-clinical evaluation exercise (mcex): Do medical students' mCEX ratings correlate with future clinical exam performance? *Academic Medicine: Journal of the Association of American Medical Colleges, 84*(10), S17–S20.

Norcini, J. J., & Blank, L. L. (1995). The mini-CEX (clinical evaluation exercise): A preliminary investigation. *Annals of Internal Medicine, 123*(10), 795–799.

Pelgrim, E. A., Kramer, A. W., Mokkink, H. G., van den Elsen, L., Grol, R. P., & van der Vleuten, C. P. (2010). In-training assessment using direct observation of single-patient encounters: A literature review. *Advances in Health Sciences Education*. doi:10.1007/s10459-010-9235-6.

Reckase, M. D. (1998). The interaction of values and validity assessment: Does a test's level of validity depend on a researcher's values? *Social Indicators Research, 45*(1/3, Validity Theory and the Methods Used in Validation: Perspectives from Social and Behavioral Sciences), 45–54.

Sidhu, R. S., Hatala, R., Barron, S., Broudo, M., Pachev, G., & Page, G. (2009). Reliability and acceptance of the mini-clinical evaluation exercise as a performance assessment of practicing physicians. *Academic Medicine: Journal of the Association of American Medical Colleges, 84* (10), S113–S115.

Sireci, S. G. (1998). The construct of content validity. In B. D. Zumbo (Ed.), *Validity theory and the methods used in validation: Perspectives from the social and behavioral sciences* (pp. 83–117). Amsterdam: Kluwer Academic Press.

Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19–37). Charlotte: Information Age Publishing.

Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice, 25*(3), 27–34.

Weller, J. M., Jones, A., Merry, A. F., Jolly, B., & Saunders, D. (2009). Investigation of trainee and specialist reactions to the mini-clinical evaluation exercise in anaesthesia: Implications for implementation. *British Journal of Anaesthesia, 103*(4), 524–530.

Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65–82). Charlotte: Information Age Publishing.