

Sources of Validity Evidence for Educational and Psychological Tests:

A Follow-Up Study

Gregory J. Cizek, University of North Carolina at Chapel Hill

Daniel Bowen, Measurement, Inc., Durham, NC

Keri Church, LEARN NC, Chapel Hill, NC

Paper Presented at the Annual Meeting of the  
National Council on Measurement in Education

Denver, CO

May 2010

## ABSTRACT

This study followed up on previous work that examined the incidence of reporting evidence based on test consequences in *Mental Measurements Yearbook*. In the present study, additional possible outlets for what has been called “consequential validity” evidence were investigated, including all articles published in the past 10 years in several applied journals devoted to educational assessment and educational policy, and all presentations at recent annual meetings for the three organizations that sponsor the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999). Consistent with previous findings, consequences of testing as a source of validity evidence is essentially non-existent in the professional literature and applied measurement and policy work. The paper concludes with implications of these findings for extending and refining current validity theory and validation practice.

## **Sources of Validity Evidence for Educational and Psychological Tests:**

### **A Follow-Up Study**

A variety of instruments is used in educational and psychological research and practice to obtain information for theory building and decision making. The data obtained from administering tests are often used as part of process that result in consequential decisions about individuals or systems including, for example, judging the effectiveness of educational and psychological interventions, awarding high school diplomas, selecting employees, issuing credentials or licenses in a profession, and countless other situations. Because such instruments are integral to the training, practice, development and research in the social sciences, the quality of the information yielded by tests is always of primary importance (Cone & Foster, 1991). Investigating and improving the quality of that information is the core interest of measurement specialists and an essential concern for test users and consumers of test information. As noted by Cone and Foster, it is imperative that those who use tests are able to “evaluate whether the data they obtain so cleverly and analyze so complexly are any good in the first place” (p. 653).

One of the characteristics evaluated—indeed, revered—when it comes to judging the quality of test information is validity. Although the concept of validity has evolved appreciably over the past 50 years (Geisinger, 1992), the primacy of validity has been consistently and enthusiastically endorsed. For example, over 40 years ago, Ebel referred to validity as “one of the major deities in the pantheon of the psychometrician” (p. 640); the current edition of the *Standards for Educational and Psychological Testing* (hereafter, *Standards*) identifies validity as “the most fundamental consideration in developing and evaluating tests” (AERA, APA, NCME, 1999, p. 9).

In contrast to the enduring esteem for the concept of validity, the concrete activity of validation has been singled out for an equally long history of rather anemic practice. Continuing to the very next

sentence of the quotation by Ebel just cited, he observed that “the good works done in [in the name of validity] are remarkably few” (1961, p. 640). More recently, Brennan has asserted that “validity theory is rich, but the practice of validation is often impoverished” (2006, p. 8).

Currently, the AERA/APA/NCME (1999) *Standards* identify five sources of validity evidence: 1) evidence based on test content; 2) evidence based on internal structure; 3) evidence based on relationships to other variables; 4) evidence based on response processes; and 5) evidence based on consequences of testing. Of these five, evidence based on the consequences of testing—or *consequential validity* as has become the shorthand term—has proven to be highly controversial and contested. Although origins of the concept can be traced back to Cronbach (1971), it is most closely associated with Messick (1989) who explicitly incorporated the notion of consequences into validity theory in his influential chapter in the third edition of *Educational Measurement* (Linn, 1989). In a 2x2 matrix, Messick presented what he referred to as four facets of validity. The matrix comprised four cells at the intersections of Test Interpretation and Test Use on one margin and Evidential and Consequential bases on the other.

While some of the facets (i.e., cells) in Messick’s matrix captured non-controversial aspects of validity (e.g., construct validity), the facet at the intersection of the consequential basis of test use has provoked lingering debate. Kane has noted that “consensus has not been achieved on what the role of consequences in validation should be” (2001, p. 328); Brennan has stated that “the most contentious topic in validity is the role of consequences” (2006, p. 8).

The controversy lingers, in part, because of seemingly orthogonal perspectives. For example, while some of those concerned about validity have expressed the opinion that “the matrix was a mistake” (Shepard, 1997, p. 6), others have asserted that “the consequences of an assessment procedure are *the first and most important consideration* in establishing the validity of the assessment” (International Reading Association and the National Council of Teachers of English, 1994, p. 17, emphasis added).

### Investigating Consequential Validity

Work presented previously at an AERA annual meeting provided additional insights into the

notion of consequential validity (see Cizek, Rosenberg, & Koons, 2007). The researchers investigated aspects of validity reflected in published measures currently used in educational and psychological tests by reviewing validity information appearing in the then-current edition of *Mental Measurements Yearbook* (Spies & Plake, 2005), a leading testing reference resource that provides synopses of validity evidence for 283 published instruments spanning a wide variety of measurement purposes, including educational achievement, ability, personality, career guidance, personnel selection, and others. That study addressed four topics: 1) the extent to which validity reports are aligned with major tenets of modern validity theory; 2) the sources of validity evidence typically reported; 3) whether sources of validity evidence differ for various kinds of tests; and 4) which sources of validity evidence are considered most important.

Results of that study indicated that validity reporting is not often provided in terms that are consonant with modern validity theory, and that the favorability of overall judgments about a test is related to the number of validity sources reported; those with more evidence were viewed more favorably, regardless of the sources of evidence provided. In addition, the authors emphasized the finding that a particular source of validity evidence—evidence based on test consequences—was essentially absent from test information and not routinely collected or reported. Specifically, whereas construct, concurrent, and content validity evidence were provided fairly frequently (in 58.0, 50.9, and 48.4% of the tests, respectively), evidence based on test consequences was noted for only two tests (0.7%). It would seem reasonable to expect that considerable attention would be given to what is deemed alternatively to be one of, or the first and most important, source of validity evidence. However, the nearly complete absence of attention to consequences suggests that consequential validity has been tacitly ignored by many, perhaps most, measurement specialists. The authors concluded that “test producers generally—and correctly—reject consequential validity” and that “the most straightforward explanation is that consequential validity simply does not exist” (Cizek, Rosenberg, & Koons, 2008, p. 410).

Three concerns arise from these findings. First, whereas the authors found that evidence about test consequences was essentially ignored, it is true that other sources of validity evidence were also not

widely reported. For example, the authors found that evidence based on response processes was provided for only 5 (1.8%) of the 283 tests. Based on that finding, it seems worth considering whether response processes as a source of validity evidence is also flawed. However, the situations differ enough that the same conclusions do not likely apply. Whereas consequences could conceivably be of concern for any test and thus would be expected to be cited much more frequently than it is at present, the same is not true for the source of evidence about response processes. That source of evidence would be germane only to the (fairly small) subset of tests that claim to measure specific constructs such as “higher order thinking,” “creativity” or other characteristics where evidence should be gathered that the test taker engages in the response processes asserted by the test developer. Because comparatively fewer educational or psychological tests make such a claim, it would be expected that fewer test reports would provide information related to this source of evidence.

A second concern is that the absence of attention to consequential validity could be interpreted as an (illogical) argument that because those who engage in validation don't do something, therefore they *shouldn't* do it. A more appropriate interpretation of the finding is that it merely helps explain *why* testing specialists have generally ignored consequences as a source of validity evidence. It is only one additional piece of information that supports a plausible hypothesis; namely, that practitioners do not appear to gather or report evidence on validity based on consequences because consequences are not a logical part of validation, and even if evidence were available about consequences, there is no way it can be synthesized with other sources of evidence to result in an overall, coherent judgment about appropriate score inferences.

On this point, it is helpful to recall that Messick defined validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences...based on test scores” (1989, p. 13). A key problem is inherent in that definition: it requires a synthetic evaluation of that which cannot be synthesized; namely, theoretical rationales and empirical evidence, along with information about social consequences. To illustrate the impossibility of the task, consider the meaning that might be made of the total score on a test comprising

20 items measuring French vocabulary and 20 items measuring Geometry. An overall index of performance could be calculated, but a meaningful interpretation of, say, a raw score of 35 on the test is impossible, and no conclusions about the examinee's standing on *either* characteristic are supportable. Analogously, evaluative judgments based on integration of the theoretical foundations underlying an instrument and information about consequences are similarly uninformative. It seems logically impossible to conceive of an integrated validity conclusion in a situation where empirical evidence showed scant support for an intended inference, but use of the test realized highly desirable social benefits.

Ultimately, evidence about the extent to which a test yields accurate inferences about a construct and evidence about the broader consequences of administering the test are not compensatory in any logical sense and cannot be combined into a coherent, integrated evaluation. Any attempt to perform an "integrated evaluative judgment" of such divergent sources of information cannot yield sound conclusions about either the scientific meaning of the performance or the desirability of using the test.

A third issue of concern is the focus of this paper. It may be somewhat of a leap to go from the analysis of a single resource (i.e., *Mental Measurements Yearbook*) to the conclusion that measurement specialists appear to have generally rejected the notion of consequential validity. Analysis of other sources of information may yield conflicting results. Thus, the purpose of the research reported in this brief note was to search other potential outlets and producers of validity evidence to see if the earlier finding was idiosyncratic or reproducible.

#### Methods

Two additional sources of information were identified as potential outlets for reports on consequential validity. First, whereas the publication of *MMY* lags somewhat behind test development and validation efforts, it might be reasoned that more timely information about consequences of testing might appear as part of reports on validation efforts in journals devoted to applied measurement. That is, it is possible that the long timeline for test development, validation, and eventual review and publication in *MMY* in 2005 may have precluded some very recent validation work involving evidence about test consequences from appearing in that source. To address this concern, we identified eight applied

measurement- and testing policy-related journals that could at least potentially publish information related to consequences of testing as a source of validity evidence. The sample of journals included on-line only and print/on-line publications, and consisted of the following: *Educational Assessment*, *Educational Evaluation and Policy Analysis*, *Educational Measurement: Issues and Practice*, *Educational Policy*, *Educational Policy Analysis Archives*, *Educational and Psychological Measurement*, *Measurement and Evaluation in Counseling and Development*, and *Practical Assessment, Research & Evaluation*. All issues of each journal for the 10-year period between the 1999 and 2008 inclusive were examined. Each article in each issue was examined for whether it focused on validity and, if so, the article was further scrutinized for mentions of information related to validity evidence based on test consequences.

Second, recalling that “evidence based on test consequences” is listed as one of the five sources of validity evidence in the *Standards*, it seemed appropriate to search for validation efforts involving consequential validity among the three sponsoring organizations of the *Standards*; namely, the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). The data source for this investigation was the annual meeting programs of each of the three sponsoring organizations. Within a given year, session titles, symposia, individual presentation titles, and keywords for each conference program were searched for the following terms: *validity*, *validation*, *consequences*, and *consequential*. At the time of data collection, the researchers had access to very recent information from each organization’s annual meeting and analyzed the two most recent years available for each organization: AERA (2007, 2008), APA (2006, 2008), and NCME (2007, 2008).

## Results

The first research question sought to identify the incidence of research or information on consequences of testing as a source of validity evidence in recent publications appearing in applied measurement and educational policy journals over the 10-year period from 1999 through 2008. Typical results from one journal, *Measurement and Evaluation in Counseling and Development*, are shown in Table 1. The columns in the table show the journal year, volume and number, total number of articles in

an issue, the number of articles dealing with validity in any respect, and the number of articles in which validity evidence based on consequences of testing is reported. As can be seen in Table 1, the most recent decade for this journal comprised 177 articles, of which 101 (57.1%) dealt in some form with the topic of validity. However, of the articles dealing with validity, no report provided information related to consequences of testing as a source of validity evidence. The results suggest that, whereas validity is often a topic of research reports in this journal, in the most recent 10-year period for which complete volumes of the journal could be examined, consideration of consequences of testing as a source of validity evidence was completely absent.

To save space, less fine-grained results are provided for the other seven journals examined (see Table 2), and the results show a continuation of the overall theme. Validity is routinely referenced throughout the applied measurement and educational policy literature; however, studies showing evidence of consequential validity are absent. As Table 2 shows, 2,408 articles were published in applied measurement and educational policy journals over the last 10 years, of which 1,007 (41.8%) dealt with the topic of validity in some way. In one article that came closest to meeting the criterion, “a validation study to obtain consequential evidence for state assessment and accountability programs is *proposed*” (Lane & Stone, 2002, p. 23, emphasis added); however, the article did not actually gather and report validity evidence based on test consequences. In fact, of the 1,007 articles reviewed, none provided evidence of consequential validity. Overall, the findings with respect to the first research question suggest that, in a sense, the virtual senate of specialists in applied measurement and educational policy voted with their work which, with remarkable consistency ignored any attention to gathering and reporting evidence of consequences of test use as a source of validity evidence.

-----  
Insert Tables 1 and 2 about here.  
-----

The second research question investigated whether the validity evidence based on consequences of testing could be found in presentations of research at the annual meetings of the organizations that

sponsor the *Standards for Educational and Psychological Testing* (AERA, APA, NMCE, 1999). It might be reasoned that applied measurement journals would have no particular interest in reporting evidence based on test consequences, but that greater sensitivity to that topic would be evident with respect to the organizations which sponsored the document that codified test consequences as a source of validity evidence.

Table 3 shows the results of analyses of conference presentations using the following keywords and search terms: *validity*, *validation*, *consequences*, and *consequential*. The results in Table 3 are presented separately for two recent, contiguous annual meetings of the three organizations. Consider the first annual meeting for each pair presented in the table; that is, the 2007 AERA annual meeting, the 2007 APA annual meeting, and the 2007 NCME annual meeting. The results for these years indicate that across the three associations, validity ( $n = 30, 47, 5$  presentations, respectively) and validation ( $n = 24, 42, 1$ , respectively) are addressed by the researchers and practitioners who participate in annual meetings. However, across the three organizations, very little attention was paid to the topics of consequences ( $n = 6, 0, 1$ ). The intersection of the terms “consequential” and “validity” was never found.

A review of all paper titles and session descriptions identified using any of the search terms was conducted to ensure that some cases of consequential validity were not overlooked. For example, it was possible that a presentation identified with the keyword “validity” might address the validity of a test use or the effects of a test. However, that was never the case. In all instances, validity research reports were clearly focused on content, predictive, concurrent, or construct validity concerns. Finally, an in-depth review of the few cases in which “consequences” or “consequential” appeared (7 and 0 instances, respectively for the first year of each annual meeting studied). This analysis revealed that the topic addressed was never the consequences of a test, but the consequences of a policy. For example, of the six AERA papers presented in 2007 with “consequences” in the title or description, three addressed the consequences of the accountability requirements of the *No Child Left Behind Act* (2001), two addressed the consequences of a specific state accountability system, and one addressed consequences of implementation of a pre-service teacher portfolios. Of the eight sessions on consequences at the 2008

AERA meeting, one paper addressed the consequences of academic ranking, three addressed student grouping practices, one focused on theory in special education, one examined the impact of inadequate education, and one investigated school choice. In summary, all of the presentations on consequences at the conference addressed systemic consequences, not those of specific tests or instruments, and none addressed consequences as a source of validity evidence.

Overall, as regards the second research question, the failure to attend to test consequences as a source of validity evidence is a consistent and persistent characteristic of the research presented at the annual meetings of the organizations that sponsor the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999).

### Conclusions and Discussion

The research reported here was conducted in an attempt to shed further light on previous findings by Cizek, Rosenberg, and Koons (2007) that showed one source of validity evidence—evidence based on consequences of testing—to be routinely ignored in the major test review resource, *Mental Measurements Yearbook*. Using other data sources, the results of the present study confirm the previous findings. In a sample of applied measurement and testing policy journals over the last 10 years including over 2400 articles many of which dealt with validity, there was not a single instance in which validity evidence based on test consequences was provided. Furthermore, in a review of two recent years of the annual meeting programs for the three professional associations sponsoring the *Standards for Educational and Psychological Testing*, attention to consequences of testing was similarly absent.

There may be possible reasons for the strikingly consistent tendency of researchers and practitioners to ignore the collection and reporting of so-called consequential validity evidence. For one, by definition, consequences cannot be observed and studied until after a test has been in operational use for some time. Thus, it may be that at least some attention to validity evidence based on consequences of testing is on-going and may be seen in future journal articles and conference presentations. On the other hand, the notion of evidence based on test consequences has been a feature of validity theory for more than two decades since the publication of Messick's (1989) influential chapter on the topic. It would seem

that sufficient time has elapsed to permit publication of validation efforts in which test consequences are reported and synthesized into coherent validity summaries.

Another possible reason for the absence of attention to the particular source of evidence known as consequential validity could be that is more difficult or more costly to gather, and that those engaged in validation efforts favor easier, cheaper sources. This hypothesis might warrant further investigation. Nonetheless, it would seem reasonable to expect that at least some attention would be given to what has been argued is a primary source of validity evidence, and the results of the present study reveal a nearly complete absence of attention to validity evidence based on consequences of testing.

This study provides support for a plausible alternative hypothesis to those just noted; namely, that researchers and practitioners do not gather or report evidence on validity based on consequences because consequences are not a logical part of validation or, as concluded in an earlier study, that consequential validity is “simply a flaw in modern validity theory” (Cizek, Rosenberg, & Koons, 2008, p. 410).

The essence of the flaw is readily discernable. Recall that Messick (1989) defined validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences...based on test scores” (1989, p. 13). For evidence based on test consequences to be a vital part of validity, it is necessary for that source of evidence to be amenable to the integration Messick described. For example, in the case of a test developed to yield inferences about “test anxiety,” it would be necessary that evidence the test was grounded in a theory of test anxiety, evidence based on test content, evidence based on correlations with tests measuring similar constructs, and evidence based on consequences of using the test be synthesized to arrive at an overall evaluative judgment regarding the degree of support for the intended inferences. However, whereas stronger evidence based on some sources of validity evidence might be viewed as compensating for comparatively weaker evidence from other sources, the case seems qualitatively different as regards what is called consequential validity evidence. It is difficult to imagine how even the most favorable evidence regarding consequences could compensate for weak evidence based on test content, relationships to other variables, response process, or internal structure to bolster confidence in the

intended score inference. As has been argued previously in this paper, even if evidence were available about consequences, there is no way it can be synthesized with other sources of evidence to result in an overall, coherent judgment about the meaning of scores—that is, about validity. Indeed—and this fact seems compelling—we have been unable to locate a single case in which an example of such a synthesis has been produced. That is, in no instance has evidence based on test consequences and any other sources of validity evidence been integrated to yield a coherent, overall case for or against the validity of scores derived from an instrument in education or psychology.

In summary, whereas a number of explanations for why a putatively vital source of validity evidence is uniformly ignored, none is as parsimonious as the proposition that consequential validity does not exist; evidence based on test consequences cannot meaningfully be incorporated into validity theory or into validation practice. That consequential validity may not exist is, perhaps, troubling and the error of including consequences in modern validity theory is not redressed simply by excising consequences from the theory or the practical guidelines that exist for best practices in test validation (e.g., the *Standards*). The consequences of testing are real, diverse, important, and warrant consideration and evaluation. They cannot be dismissed by ignoring them and they must be accounted for in a comprehensive approach to defensible test use.

Given the results of this study, there would now seem to be fairly abundant evidence to suggest that further documentation of the absence of attention to consequences of testing as a source of validity evidence is unnecessary. What is needed as a next step is a differentiation between validation practice (that is, the methods and sources of information relevant to determining the confidence that is warranted regarding intended test score inferences) and justification of test use (that is, the methods and sources of information—including consequences—brought to bear on the question of whether it is a good idea to use a given test in the first place.) As is perhaps obvious, strongly supportive evidence regarding the former is a necessary but not sufficient condition for the latter. If, indeed, efforts to establish the validity of test score inferences can be seen as distinct from efforts to justify a particular test use, then clearly much work lies ahead, not only in fostering improved good works related to validity, but in developing a nascent

theory and best practices of information gathering and evaluation of the evidence to support test use.

## References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.), *Educational measurement, fourth edition* (pp. 1-16). Westport, CT: Praeger.
- Cizek, G. J., Rosenberg, S., & Koons, H. (2007, April). *Sources of validity evidence for educational and psychological tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Cizek, G. J., Rosenberg, S., & Koons, H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68, 397-412.
- Cone, J. D., & Foster, S. L. (1991). Training in measurement: Always the bridesmaid. *American Psychologist*, 46, 653-654.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement, 2<sup>nd</sup> ed.* (pp. 443-507). Washington, DC: American Council on Education.
- Ebel, R. L. (1961). Must all tests be valid? *American Psychologist*, 16, 640-647.
- Geisinger, K. F. (1992). The metamorphosis of test validation. *Educational Psychologist*, 27, 197-222.
- International Reading Association & National Council of Teachers of English. (1994). *Standards for the assessment of reading and writing*. Newark, DE: International Reading Association.
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23-30.
- Linn, R. L. (Ed.). (1989). *Educational measurement, third edition*. New York: Macmillan.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement, 3<sup>rd</sup> ed.* (pp. 13-103). New York: Macmillan.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational*

*Measurement: Issues and Practice*, 16(2), 5-8, 13, 24.

Spies, R. A., & Plake, B. S. (Eds.). (2005). *The sixteenth mental measurements yearbook*. Lincoln, NE:

Buros Institute of Mental Measurements.

Table 1

Consequences of Testing as Validity Evidence in *Measurement and Evaluation in Counseling and Development*, 1999-2008

Year	Volume (Number)	Total Number of Articles	Number of Validity Articles	Articles Including Evidence on Consequences
1999	31(4)	4	1	0
	32(1)	4	1	0
	32(2)	5	2	0
	32(3)	4	2	0
2000	32(4)	4	2	0
	33(1)	4	2	0
	33(2)	4	1	0
	33(3)	4	2	0
2001	33(4)	5	2	0
	34(1)	4	3	0
	34(2)	4	4	0
	34(3)	5	2	0
2002	34(4)	4	3	0
	35(1)	5	3	0
	35(2)	5	1	0
	35(3)	6	0	0
2003	35(4)	6	3	0
	36(1)	5	0	0
	36(2)	5	3	0
	36(3)	5	3	0
2004	36(4)	4	4	0
	37(1)	4	3	0
	37(2)	5	2	0
	37(3)	5	2	0
2005	37(4)	5	3	0
	38(1)	4	2	0
	38(2)	5	5	0
	38(3)	5	4	0
2006	38(4)	6	3	0
	39(1)	3	2	0
	39(2)	4	4	0
	39(3)	4	2	0
2007	39(4)	4	4	0
	40(1)	4	4	0
	40(2)	4	4	0
	40(3)	5	3	0
2008	40(4)	3	3	0
	41(1)	5	4	0
	41(2)	3	1	0
	41(3)	3	2	0
Totals (%)		177 (100.0)	101 (57.1)	0 (0.0)

Table 2

Consequences of Testing as Validity Evidence in Applied Measurement and Testing Policy-Related Journals, 1999-2008

<b>Journal Name</b>	<b>Total Number of Articles, 1999-2008</b>	<b>Number of Validity Articles</b>	<b>Articles Including Consequences As Validity Evidence</b>
<i>Educational Assessment</i>	109	51	0
<i>Educational Evaluation and Policy Analysis</i>	189	73	0
<i>Educational Measurement: Issues and Practice</i>	323	61	0
<i>Educational Policy</i>	376	80	0
<i>Educational Policy Analysis Archives</i>	441	68	0
<i>Educational and Psychological Measurement</i>	642	486	0
<i>Measurement and Evaluation in Counseling and Development</i>	177	101	0
<i>Practical Assessment, Research &amp; Evaluation</i>	151	105	0
<b>Totals</b>	2408	1007	0

Table 3

Incidence of Consequential Validity in Professional Association Presentations by Year

	Professional Association						Totals
	AERA		APA		NCME		
Search Term	2007	2008	2006	2008	2007	2008	
<i>validity</i>	30	28	47	36	5	7	153
<i>validation</i>	24	32	42	26	1	1	126
<i>consequences</i>	6	23	0	16	1	1	47
<i>consequential</i>	0	0	0	0	0	0	0
Totals	54	83	89	78	7	9	326