**VALIDITY AND VALIDATION IN ASSESSMENT:**

**SELECTED REFERENCES**

**(last updated 15 June 2014)**

Alderson, J. C. (1988). New procedures for validating proficiency tests of ESP? Theory and practice.*Language Testing, 5*(2), 220-232.

Allison, D., & Cheung, E. (1991). ‘Good’ and ‘poor’ writing and writers: Studying individual performance as a part of placement test validation. *Hong Kong Papers in Linguistics and Language Teaching, 14,* 1-14.

Arkoudis, S., & O’Loughlin, K. (2004). Tensions between validity and outcomes: Teachers’ assessment of written work of recently arrived immigrant ESL students. *Language Testing, 20*, 284-304.

Aryadoust, V. (2013). *Building a validity argument for a listening test of academic proficiency*. Cambridge, UK: Cambridge Scholars Publishing.

Bachman, L. F. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16*(4), 449-465.

Bachman, L. F. (1988). Problems in examining the validity of the oral proficiency interview. *Studies in Second Language Acquisition, 10*, 149-164.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford, UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1981). The construct of validation of the FSI oral interview. *Language Learning, 31*, 167-186.

Bachman, L. F., & Palmer, A. S. (1981). A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence,* (pp. 149-165). Washington, DC: TESOL Publications.

Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly, 16*, 449-465.

Banerjee, J., & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham & D. Corson (Eds.), *Language testing and assessment. Encyclopedia of Language and Education* (Vol. 7, pp. 275-287). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bateman, H. (2010). A study of the context and cognitive validity of a BEC vantage test of writing. *Cambridge ESOL Research Notes*, *42*, 40.

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test.*Language Testing, 27*, 101-118.

Bennett, R. E. (2004). *Moving the field forward: Some thoughts on validity and automated scoring*. Princeton, NJ: Lawrence Erlbaum Associates.

Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It’s not only the scoring. *Educational Measurement: Issues and Practice, 17*(4), 9-17.

Benson, J., Moulin-Joulin, M., Schwarzer, C., Seipp, B. & El-Zahhar, N. (1992). Cross validation of a revised test anxiety scale using multi-national sample. In K.A. Hagtver & T.B. Johnson (Eds.), *Advances in test anxiety research* (pp. 62-83). Amsterdam, the Netherlands: Swette & Zeitlinger.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.

Breeze, R., & Miller, P. (2012) Predictive validity of the IELTS listening test as an indicator of student coping ability in English-medium undergraduate courses in Spain. In L. Taylor & C. Weir (Eds.), *Studies in Language Testing 34: Research in reading and listening assessment* (pp. 487-518). Cambridge, UK: Cambridge University Press.

Brennan, R. L. (Ed.). (2006). *Educational measurement, 4th Ed.* Washington, DC: American Council on Education.

Bridges, G. (2010). Demonstrating cognitive validity of IELTS academic writing task 1. *Cambridge ESOL Research Notes*, *42,* 24-33.

Brown, J. D. (2005). Language test validity. *Testing in language programs: A comprehensive guide to English language assessment* (pp. 220-251). New York, NY: McGraw-Hill.

Brown, J. D., Cunha, M. I. A., & Frota, S. (2001). The development and validation of a Portuguese version of the motivated strategies for learning questionnaire. In Z. Dornyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 257-280). Honolulu, HI: University of Hawaii Press.

Camp, R. (1993). Changing the model for the direct assessment of writing. In M. M. Williamson & B. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. (pp. 45–78). Cresskill, NJ: Hampton Press.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Thousand Oaks, CA: Sage Publications, Inc.

Castro, S., & Lima, C. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody.*Behavior Research Methods, 42*(1), 74-81. Retrieved from <http://ulib.iupui.edu/cgi-bin/proxy.pl?url=/docview/204304709?accountid=7398>

Chapelle, C. (1998) Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (Eds.), *Second language acquisition and language testing interfaces* (pp. 32-70). Cambridge, UK: Cambridge University Press.

Chappelle, C. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, *19*, 254-272. doi:10.1017/S0267190599190135

Chapelle, C. A., Enright, M. & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language™.* London, UK: Routledge.

Clark, J. L. D. (1988). Validation of a tape-mediated ACTFL/ILR-scale based test of Chinese speaking proficiency. *Language Testing, 5*, 187-205.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cronback, L. J. (1989). Construct validity after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 147-171). Urbana, IL: University of Illinois Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302

Cumming, A. (1996). Introduction: The concept of validation in language testing. In A. Cumming & R. Berwick (Eds.), *Validation in Language Testing* (pp. 1-14). Clevedon, UK: Multilingual Matters Ltd.

Cumming, A., & Berwick, R. (Eds.). (1996). *Validation in language testing.* Clevedon, UK: Multilingual Matters Ltd.

Cumming, A., & Mellow, D. (1996). An investigation into the validity of written indicators of second language proficiency. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 72-93). Clevedon, UK: Multilingual Matters.

Cushing Weigle, S., & Lynch, B. (1996). Hypothesis testing in construct validation. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Dahllöf, U. S. (1971). *Ability grouping, content validity and curriculum process analysis.* New York, NY: Teachers College Press.

Dandonoli, P., & Henning, G. (1990). An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals, 23*, 11-22.

Davies, A. (1996). The role of the segmental dictionary in professional validation: Constructing a dictionary of language testing. In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 222-235). Clevedon, UK: Multilingual Matters.

Davies, A., & Elder, C. (2011). Validity and validation in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 705-813). New York, NY: Routledge.

Davis, K. A. (1992). Validity and reliability in qualitative research on second language acquisition and teaching. *TESOL Quarterly, 26*, 605-608.

Elder, C., & Wigglesworth, G. (2006). An investigation of the effectiveness and validity of planning time in Part 2 of the IELTS Speaking Test. In P. McGovern & S. Walsh (Eds.), *IELT Research reports Volume 6* (pp. 13-40). Canberra, Australia: IELTS Australia and the British Council.

Elliott, M. & Wilson, J. (2011). Context validity. In L. Taylor (Ed.), *Studies in language testing, 30:* *Examining speaking: Research and practice in assessing second language speaking* (pp. 152-241). Cambridge, UK: UCLES/Cambridge University Press.

Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Studies in language testing, 30:* *Examining speaking: Research and practice in assessing second language speaking* (pp. 65–111). Cambridge, UK: UCLES/Cambridge University Press.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary.*Language Testing, 27*, 537-554.

Frederiksen, N. (1986). Construct validity and construct similarity: Methods for use in test development and test validation. *Multivariate Behavioral Research*, *21*(1), 3-28.

Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied Linguistics, 20*, 221-236.

Gellert, A., & Carsten, E. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment, 31*(1), 16-28.

Geranpayeh, A. (2011). Scoring validity. In L. Taylor (Ed.), *Studies in language testing, 30:* *Examining speaking: Research and practice in assessing second language speaking* (pp. 242-272). Cambridge, UK: UCLES/Cambridge University Press.

Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing, 3*, 159-185.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23,* 17-27.

Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on validity: A historical analysis of language testing conference abstracts. In A. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium* (pp. 253-276). Mahwah, NJ: Lawrence Erlbaum Associates.

Hawkey, R. (2011). Consequential validity. In L. Taylor (Ed.), *Studies in language testing, 30:* *Examining speaking: Research and practice in assessing second language speaking* (pp. 273-302). Cambridge, UK: UCLES/Cambridge University Press.

He, L. Z., &Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing 23*, 370-401.

Henning, G. (1983). Oral proficiency testing: Comparative validities of interview, imitation, and completion methods. *Language Learning, 33,* 315-332.

Henning, G. (1992). The ACTFL oral proficiency interview: Validity evidence. *SYSTEM, 20*, 365-372.

Hill, K. (1997). The role of questionnaire feedback in the validation of the oral interaction module. In G. Brindley & G. Wigglesworth (Eds.), *Access: Issues in language test design and delivery* (pp. 147-174). Sydney, Australia: NCELTR.

Holliday, A. (2004). Issues in validity in progressive paradigms of qualitative research. *TESOL Quarterly, 38*(4), 731-734.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review, 64*(4), 555-580.

Jafarpur, A. (1996). Native speaker performance validity: In vain or for gain? *System, 24*(1), 83-95.

Jin, Y. & Cheng, L. (2013). The effects of psychological factors on the validity of high-stakes test. *Modern Foreign Languages (Quarterly), 36*(1), 62-69.

Johnson, D. M., & Saville-Troike, M. (1992). Validity and reliability in qualitative research on second language acquisition and teaching. *TESOL Quarterly, 26*, 602-605.

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices, 21*(1), 31-41.

Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research & Perspective, 2(3)*, 135-170.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Washington, DC: National Council on Measurement in Education & American Council on Education.

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing, 29*, 3-17.

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

Kane, M. T. (2013). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement*, *50*(1), 115-122. doi:10.1111/jedm.12007

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*(2), 5-17.

Khalifa, H. (2010). Construct validation of the reading module of an EAP proficiency test battery. *Cambridge ESOL Research Notes*, *42*, 8-14.

Kunnan, A. (1998). Approaches to validation in language assessment. In A. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium, Long Beach* (pp. 1-16). Mahwah, NJ: Lawrence Erlbaum Associates.

Kunnan, A. (1998). Preface. In A. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium, Long Beach* (pp.ix-x). Mahwah, NJ: Lawrence Erlbaum Associates.

Kunnan, A. (Ed.). (1998). *Validation in language assessment: Selected papers from the 17th language testing research colloquium, Long Beach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kunnan, A. J. (2000). *Fairness and validation in language assessment.* Cambridge, UK: Cambridge University.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests.* Cambridge, UK: Cambridge University Press.

Lennon, R. T. (1956). Assumptions underlying the use of content validity. *Educational and Psychological Measurement, 16*, 294-304.

Lim, G. (2013). Components of an elaborated approach to test validation. *Cambridge English Research Notes, 51*, 11-14.

Lim, G. S. (2012). Developing and validating a mark scheme for writing. *Research Notes, 49*, 6-1.

Lim, G.S. & Khalifa, H. (2011). Criterion-related validity. In L. Taylor (Ed.), *Studies in language testing, 30:* *Examining speaking: Research and practice in assessing second language speaking* (pp. 303-321). Cambridge, UK: UCLES/Cambridge University Press.

Lin, C. K., & Zhang, J. (2013). Enhancing standard-based validity for ELL population: A perspective from correspondence between standards. *TESOL Quarterly, 47*(2), 399-410.

Lindquist, E. F. (Ed.). (1951). *Educational measurement.* Washington, DC: American Council on Education.

Linn, R. L. (Ed.). (1989). *Educational measurement, 3rd Ed*. Washington, DC: American Council on Education.

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice, 16*, 28-30.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 15-21.

Long, M. H. (1997). Construct validity in SLA research. *The Modern Language Journal, 81,* 318-23.

Low, G. (1985). Validity and the problem of direct language proficiency tests. In J. C. Alderson (Ed.), *Lancaster papers in English language education:* *Evaluation* (pp. 151-168). Oxford, UK: Pergamon Press.

Lowe, P. (1981). Structure of the oral interview and content validity. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence* (pp. 71-80). Washington, DC: TESOL.

Lyons, N. (Ed.). (1998). *With portfolio in hand: Validating the new teacher professionalism.* New York, NY: Teachers College Press.

Markus, K. A. (1998). Measurement and validity: Is completion of Samuel Messick’s synthesis possible? *Social Indicators Research, 45*(1/3), 7-34.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 25*, 1012-1027.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning of consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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

Messick, S. (1996). Standards-based score interpretation: Establishing valid grounds for valid inferences. *Proceedings of the joint conference on standard setting for large scale assessments,* Sponsored by National Assessment Governing Board and The National Center for Education Statistics. Washington, DC: Government Printing Office.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*, 241-256.

Messick, S. (1996). Validity of performance assessment. In G. W. Philips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1-18). Washington, DC: National Center for Educational Statistics.

Mickan, P., Slater, S., & Gibson, C. (2000). Study of response validity of the IELTS Writing Subtest. In R. Tulloh (Ed.), *IELTS research reports,* *Vol. 3* (pp. 29-48). Canberra, Australia: IELTS Australia.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*(4), 13-25.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30*, 109-162.

Palmer, A. S., Groot, P. J. M., & Trosper, G. A. (Eds.). (1981). *The construct validation of tests of communicative competence.* Washington, DC: TESOL.

Pearlman, M. (2008). Finalizing the test blueprint. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 227- 258). New York, NY: Routledge.

Pomerantz, A. (1990). Chautauqua: On the validity and generalizability of conversational analysis methods: Conversation analytic claims. *Communication Monographs, 57*, 231-235.

Reckase, M. (1998). Consequential validity from the test developer’s perspective. *Educational Measurement: Issues and Practice, 17*, 13-16.

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing, 23,* 229-256.

Rose, K. R. (1994). On the validity of discourse completion tests in non-Western contexts. *Applied Linguistics, 15*, 1-14.

Sakyi, A. (2000). Validation of holistic scoring for writing assessment: How raters evaluate ESL compositions. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 129-152). Cambridge, UK: Cambridge University Press.

Schrauf, R. W., Weintraub, S., & Navarro, E. (2006). Is a validation of the Word Accentuation Test (WAT) of premorbid intelligence necessary for use among older, Spanish-speaking immigrants in the United States?  Journal of the International Neuropsychological Society, 12, 391-399.

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.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shohamy, E. (1994). The role of language tests in the construction and validation of second-language acquisition theories. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 133-142). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11,* 99-123.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *Modern Language Journal, 95*(1), 418-429.

Stevenson, D. K. (1981). Beyond faith and face validity: The multitrait-multimethod matrix and the convergent and discriminant validity of oral proficiency tests. In A. S. Palmer, P. J. M. Groot, & G. A. Trosper (Eds.), *The construct validation of tests of communicative competence* (pp. 37-61). Washington, DC: TESOL.

Thorndike, R. L. (Ed.). (1971*). Educational measurement, 2nd Ed.*  Washington, DC: American Council on Education

Underhill, N. (1983). Commonsense in oral testing: Reliability, validity and affective factors. In M. A. Clarke, & J. Handscombe (Eds.), *On TESOL ’82: Pacific perspectives on language learning and teaching* (pp. 125-139). Alexandria, VA: TESOL.

Urmston, A., **Raquel, M., & Tsang, C.** (2103). **Diagnostic testing of Hong Kong tertiary students’ English language proficiency: The development and validation of DELTA.** *Hong Kong Journal of Applied Linguistics 14*(2), 60-82.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing, 23*(4), 411-440.

Van Moere, A. (2007). Validity evidence in a group oral test. *Language Testing, 23*, 411-440.

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*, 335-353.

Weinrott, M. R., Jones, R. R., & Bolder, G. R. (1981). Convergent and discriminant validity of five classroom observation systems: A secondary analysis. *Journal of Educational Psychology, 73*(5), 671-680.

Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, UK: Palgrave Macmillan.

Williamson, M. M., & Huot, B. (Eds.), (1993). *Validating holistic scoring for writing assessment: Theoretical and empirical foundations*. Cresskill, NJ: Hampton Press.

Wilson, K. M., & Graves, K. (1999). *Validity of the secondary level English proficiency test at Temple University Japan* (ETS RR-99-11). Princeton, NJ: ETS.

Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers’ perceptions matter. *TESOL Quarterly, 45*(4), 628-660.

Xie, Q. (2011). Is test taker perception of assessment related to construct validity? *International Journal of Testing, 11*(4), 324-348.