English as Second Language: Alignment Perspective as Quality Improvement in Pakistan

Abida Khalid
Assistant Professor
University of Education
Bank Road Campus, Lahore, Pakistan

Muhammad Azeem
Assessment Expert
Punjab Education Assessment System (PEAS)
University of Education
College Road, Township, Lahore, Pakistan

Dr Muhammad Bashir Gondal
Senior Research Fellow
Punjab Education Commission (PEC)
Wahdat Colony Road, DSD, Lahore, Pakistan

Abstract

Traditional method for the assessment of students' writing skills in English Language, involves constructing a test, directly from content area, and mostly without blueprint. A test blueprint usually specifies the extent of content within the tests only while alignments often overlooked. Inexperienced—non expert in assessment test developer—especially untrained teachers, rate the students on their own (subjective) point of view irrespective of demand of curriculum objectives and market. In result our young talent, due to limitations of assessment system, is not streaming with national and global requirements. Broadly speaking, alignment, in the context of education, can be defined as the degree to which the components of an education system—such as standards, curricula, assessments, and instruction—work together to achieve desired goals. Systematic review of standards and items can yield judgments related to broad objective coverage, range of coverage, balance of coverage, and depth coverage. These four—Knowledge Concurrence, range of knowledge, balance of representation, and depth of knowledge are the key features of Webb's Model. Thus Webb's model, in Pakistan, can be implemented for the alignment of standards of English language, curricula and assessment for precise interpretation of students' attainments and proficiencies in the subject of English as a second language.

Introduction

A typical method for creating tests, in all subjects, based on objectives, involves constructing a test "blueprint" for each grade and content area. A test blueprint usually specifies the extent of content within the tests while item specification specifies the range of difficulty for test items, and the structure of the test. This blueprint should be linked directly with content standards. Most commonly, test blueprints are arranged around content areas by noting each standard and objective that is tapped by this content goal. Often, an approximately equal number of test items are constructed corresponding with each content area of the blueprint. Alternatively, a test may be "weighted" with more items included for a particular content area due to the organization of the standards and test blueprints. For example, a state may expect students to develop and demonstrate a broader range of English language skills or a more specific portion of mathematics, and these skills would be represented by more test items. The continuity between the content standards, test blueprints, and actual test items is crucial. All three documents should match very well. Inconsistency between any of these documents can lead to "poor alignment," in which case the results from the assessment are not as meaningful.

The Concept of Alignment

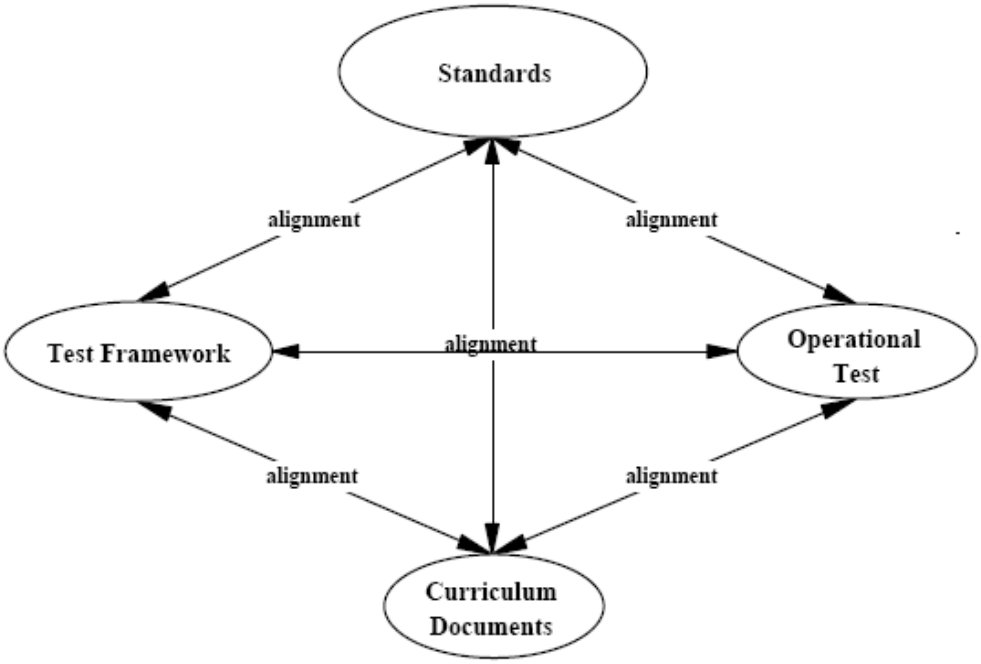
"Alignment is central to current efforts of systemic and standards-based education reforms in mathematics and science (Webb 1997)". Various countries have content frameworks for different subjects and art measuring students' attainment of expectations given in the frameworks through large scale assessments. Educational policies are based on expectations—what students should know and what they should be able to do with that knowledge while assessments—measuring tools are being used to measure students' achievement by teachers.

All assessments used constitute an assessment system—first standards, then curriculum frameworks, and then assessments. Alignment is the extent to which "expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do" (Webb et al., 2002, p. 1) Alignment is the measure of the degree of match between content included in test and the content of subject area through national academic standards.

Alignment: Standard and Assessment

As cited by Betsy, Margaret and Zucker (2004) "In the context of education, alignment can be broadly defined as the degree to which the components of an education system—such as standards, curricula, assessments, and instruction—work together to achieve desired goals (Ananda, 2003; Resnick, Rothman, Slattery, and Vranek, 2003; Webb, 1997b)". Various alignment studies, all over the world, are being examined the degree to which content standards and content of assessment are matched. Educational researchers are extending alignment to standards, assessments, and curriculum and instruction. Resulting significance of accountability in education, the attention of education policymakers and researchers has recently turned to alignment. Alignment conceptions have become increasingly mandatory to meet goals for strengthening education systems and to satisfy rigorous requirements for accountability assessments (Betsy, 2004).

Alignment across Documents



Derived from An Agenda for NAEP Validity Research, 2002

Marca (2001) states "To make defensible accountability decisions based in part on students' academic achievement and proficiencies, states must employ assessments that are aligned to their academic standards". It is just one of several criteria for test/ assessment validation process Marca (2001). Alignments have significant role in methodological and ethical requirement. As methodological requirement classroom instruction, standards, framework, and test should be align and for proficiency testing content standards, performance standards, framework and assessments while for ethical requirements test score should reflect performance relative to skills that adequately represent our expectations for required achievement.

In Pakistan there is lack of standard-based assessment in the subject of English Language. English Language skills of students are being assessed by achievement tests only. Test developers rarely take in to account the alignment of Test and class-room Instruction. This leads towards invalid and low quality of assessment of students skills in ESL. The concept of alignment for valid assessment and quality insurance is emerging field in Education in technologically advanced countries. Pakistan may compete globally by adopting alignment methodologies.

Methods for Alignment

Commonly following three methods for systematically evaluating and documenting the alignment between standards and assessments of English as Second Language (ESL) are in use:

- sequential development
- expert review
- and document analysis

Sequential Development

It is easiest alignment method. In this method first, educators, experts, and the society are designed the subject standards. Second, these standards are using for the structure blueprint and content of the assessment. So this method establish link between each standard and each assessment item for evidence of alignment.

Expert Review

In this method first, both standards and assessments are developed. After the development of standards and assessments this method analyzes the alignment between them. Various technical groups—educators, Item developers, curriculum experts, and standards developers, check the alignment within in the framework by discussion and with the help of teachers. Then a panel of test development experts—knowledgeable about the content covered by the standards and about the process for developing tests, compares the standards to the assessment. This method is an extended form of sequential development. Expert review provides evidence of alignment between standards and an assessment.

Document Analysis

This is a complicated alignment method. In this method first, content and structure of the standards and assessment documents are encoded and then are analyzed. The alignment of the documents can then be quantified and systematically compared. This methodology is being used by Third International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA)

These three alignment methodologies have been used successfully for the subjects of English Language and Mathematics in various education systems especially in United States. Moreover, education research continues to strengthen the understanding and practical application of these methodologies. The methodologies can be used independently or in a combination of the three. In Pakistan to align newly k-12 developed standards for different subjects and accountability assessments, policymakers may select the combination of sequential development, and expert review that is suited to our education system's needs and resources. For more accurate alignment of assessment with academic content standards, teachers and policymakers will have workout for getting accurate information about students' achievement and proficiencies. With this information, educators and teaches may be able to improve instruction in the classroom resultantly policymakers will have information to improve the education system as a whole.

Models of Alignment

Following alignment models are in use for different subjects especially for English Language. These alignment models are supported by the Council of Chief State School Officers (CCSSO), an organization that assists States in measuring and meeting achievement goals.

Webb Model

This is the most popular alignment model. It provides analysis of the degree of intersection between assessments and content standards, which combines qualitative expert judgments and quantified by coding and statistical analysis. This model has been in practice in various states in USA. Norman Webb, with assistance from CCSSO, was developed this model.

Surveys of Enacted Curriculum (SEC) Model

It produces alignment analyses of standards, assessments, and instructional content by use of a content matrix and allows comparison across different schools, districts. Andrew Porter, Director of the Wisconsin Center for Education Research, and John Smithson, with assistance from CCSSO, was developed this model.

Achieve

Achieve, Inc., a nonprofit education leadership organization based in Washington, DC has developed the Achieve model.

CBE

The Council for Basic Education (CBE), based in Washington, DC, is providing technical support for alignment analysis.

Key features of Alignment Models

These four models are in use but their application is based on resources. Webb's model is less complex than other three methods. The big advantage is that it is independent of coding by reviewers and statistical procedures are used for measuring inter-rater reliability, and variation in alignment statistics. Various studies show that no method, due their limitations, can make alignment perfectly. The methods used by Webb and Porter can be implemented either by experienced educational research companies or by trained State personnel. Achieve Inc. offers product services on alignment.

Methodological Consideration

Alignment is a critical step in validation of test score. A comprehensive method certainly required in establishing evidence of the validity of test score or performance interpretation.

Various methodologies of alignment are available but, basically, the analysis of alignment requires a two-step process:

- systematic review of standards
- systematic review of benchmarks
- systematic review of indicators
- systematic review of framework
- systematic review of test items/tasks.

These processes become critical when considering the judgment of alignment in-depth.
Personals having expertise in both content area of the subject and assessment should conduct the review of standards and assessments.

Systematic review of standards/benchmarks/indicators

- How standards can be measured?
- What are the minimum expectations?
- What types of items are required to achieve standards?

Review of Framework

- Dimensions of framework
- What are content standards?
- What are performance standards?
- What skills are demanded by framework to be measured?

Review of test items/tasks

This will involve two decision points

- What objective(s) an item measures? and
- Item's degree of cognitive complexity.

Systematic review of standards and items can yield judgments related to broad standard coverage, Systematic review of benchmarks and indicators yield judgment related to range of coverage, balance of coverage, and depth coverage. There is no hard and fast rule for each alignment dimension but Webb (1999) has provided a set of decision rules for alignment analysis that should be supported by evidence of score reliability.

Significance of alignment as Quality Improvement

Messick (1989) has given new concept of validity. He argued that validity is not a quality of an assessment/test but it is related to the inferences drawn from test scores or performance. This concept of validity has totally changed the traditional conceptions of validity that focus from establishing different sorts of validity—content validity, and construct validity for the validation of test score inferences. Messick (1989) states that "Content validity is based on professional judgments about the relevance of the test content to the content of a particular behavioural domain of interest and about the representativeness with which item or task content covers that domain" (p. 17).

Evidence of test relevance and the target domain relevance is a critical first step in validating interpretations of test score. As cited by Marca and Paul (2001) "the establishment of evidence of content representativeness or alignment is intricately tied to evidence of construct validity. Although constructs are typically considered latent causal variables, their validation is often captured in measures of internal and external structure (Messick, 1989). Arguably the interpretation of measures of internal consistency and/or factor structures, as well as associations with external criterion, will be informed by an analysis of range of content and balance of content coverage".

Pakistan: Validity and Alignment

Unfortunately, in Pakistan, validity and reliability are reporting in traditional sense—face validity, construct validity, concurrent validity and correlational reliability etc while concept of alignment has no entrenched. At the time, a few studies at national level by National Educational Assessment System (NEAS) and attached provincial institutions are being conducting large-scale assessments using some extent of alignment analysis but modern concept of validity is not well established.

The reason may be there are no test specialists with advance knowledge of alignment and lack of research and its applications in Pakistani context. In advance countries a series of assessment related researches and their applications are in operations. The big question is that, in Pakistan, are we assessing our young's skills in English as Second Language for intended outcomes and resultantly what are their consequences?

The question may be answered by developing assessment tools exploring alignment between standards, instruction, subject matter of English Language—content coverage, depth of knowledge, and balance representation and validity evidence—construct validity, usability, value implications of score meaning as a basis for action and the social consequences of score use. We can make assessment more meaningful and worthy by minimizing the effects of 1) "construct underrepresentation—the tasks which are measured in the assessment fail to include important dimensions or facets of the construct and therefore, the test results are unlikely to reveal a student's true abilities within the construct which was indicated as having been measured by the test and 2) construct-irrelevant variance—the test measures too many variables, many of which are irrelevant to the interpreted construct. This type of invalidity can take two forms, "construct-irrelevant easiness" and "construct--irrelevant difficulty." "Construct-irrelevant easiness" occurs when extraneous clues in item or task formats permit some individuals to respond correctly or appropriately in ways that are irrelevant to the construct being assessed; "construct-irrelevant difficulty" occurs when extraneous aspects of the task make the task irrelevantly difficult for some individuals or groups. While the first type of construct irrelevant variance causes one to score higher than one would under normal circumstances, the latter causes a notably lower score (Amy 1999)".

Summary

The new concept of alignment is providing evidences of quality and validity of assessment of all subjects especially English Language, Science, and Mathematics. The challenge that is being faced, in Pakistan, by English Language test developers and alignment analyzers is what constitutes "good" validity evidence using different techniques. Alignment studies can improve the quality of assessing students' skills in English as Second Language. At the moment, there are no recommend minimum acceptable values of statistical indicators for all of the sources of validity evidence but alignment analysis are providing these evidences empirically. In practice, to reach full agreement between expectations and assessments on all criteria is extremely difficult. Tradeoffs will need to be made because real constraints exist on any education system such as resources, finances, time, and legal authority. Decisions on what tradeoffs are to be made among these criteria or on what level of compliance will be acceptable should be made in full awareness of potential consequences.

References

Amy, B.1999. Traditional and Modern Concepts of Validity. Washington DC: ERIC Clearinghouse on Assessment and Evaluation Publisher. (Retrieved on May 7, 2007 fromhttp://www.ericdigests.org/2000-3/validity.htm ERIC Document Reproduction Service No. ED435714)

An Agenda for NAEP Validity Research. Retrieved on 27-02-2007 from http://www.air.org/publications/documents/whitepapers.pdf

Betsy, J.C., Margaret, A. J, Zucker, S. 2004. Alignment in Educational Assessment San Antonio, TX: Harcourt Assessment, Inc.

David et al. 2006. Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application. The American Journal of Medicine (2006) 119, 166.e7-166.e16

Fulcher, G. & Davison, F. 2007. Language Testing and Assessment: An Advanced Resource Book. UK: Routledge
http://harcourtassessment.com/NR/rdonlyres/A37E5299-6483-4021-BB8A-CB10C6F89BB6/0/AlignmentMethodologies.pdf
http://harcourtassessment.com/NR/rdonlyres/C19A3D92-B124-4098-BC2F-0BCD490DCAA2/0/AlignEdAss.pdf

La Marca, Paul M. 2001. Alignment of standards and assessments as an accountability criterion. Practical Assessment, Research & Evaluation, 7(21). Retrieved May 3, 2007 from http://PAREonline.net/getvn.asp?v=7&n=21

Linn, R. L. & Gronlund, N. E. 2005. Measurement and assessment in Teaching (8th ed.). New Delhi: Baba Barkha Nath Printers

Messick, S. 1989. Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed.) (pp. 13-103). New York: American Council on Education & Macmillan.

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

Millman, J. & Greene, J. 1993. The Specification and Development of Tests of Achievement and Abilities. In R. L. Linn (Editor), Educational Measurement (3rd Edition). New York: American Council on Education – Macmillan Publishing Company.

Minnesota Department of Education. (2004, March 15). The Minnesota test of emerging academic English technical report, fall 2003 test administration.

Minnesota Department of Education. (2003, July 10). Minnesota English language proficiency standards for English language learners K-12.

Minnesota Department of Education. (2003, July 9). An alignment study of the test of emerging academic English (TEAE) and Minnesota's grade-specific expectations for reading, grades 3, 5, 7 and 10.

Nevada Department of Education. 2001. Nevada English Language Arts: Content Standards for Kindergarten and Grades 1, 2, 3, 4, 5, 6, 7, 8 and 12.

Schmidt, W. 1999. Presentation in R. Blank (Moderator), The Alignment of Standards and Assessments. Annual National Conference on Large-Scale Assessment, Snowbird, UT.

Webb, N. L. 1997. Research Monograph No. 6: Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Washington, DC: Council of Chief State School Officers.

Webb, N. L. 1999. Alignment of Science and Mathematics Standards and Assessments in Four States.

Washington, DC: Council of Chief State School Officers.

Authors' Notes

Muhammad Azeem

Mr. Muhammad Azeem is working as Assessment Expert and coordinator for District Wide Large Scale Assessment (D-WLSA) at Punjab Education Assessment System, Govt. of The Punjab, Lahore, Pakistan. He has submitted his Doctoral Thesis at the Division of Education, University of Education, Lahore under Indigenous PhD Fellowship from HEC Pakistan. His area of interest is Educational Assessment, Evaluation, and Testing. He has written a number of articles and research papers in different local and international magazines and journals. He has visited School of Education, University of Leicester, UK as Doctoral Research Fellow. He can be reached at **Knowledge_jhumra@yahoo.com**

Abida Khalid

Ms. Abida is an eminent educationist. She is working as Assistant Professor at University of Education, Bank Road campus, Lahore. She has 22 years of teaching experience at College and University level. She is a researcher of high caliber and has supervised approximately 250 Research Projects at University level. Currently, she holds the designation of Assistant Professor at University Of Education, Lahore she is Ph.D Scholar in University of management & Technology. She has compiled books on diversified topics. The titles of the books are as follows: Contemporary Modes of Thoughts in Philosophy of 21st Century, Basics of Educational Research with SPSS Analysis, Phonetics and Phonology in English, Educational Research and electronic data analysis, and Introduction to Education. Her various modes of international papers has been published.

Dr Muhammad Bashir Gondal

Dr Muhammad Bashir Gondal is former director of Punjab Education Assessment System (PEAS) and presently working as Senior Research Fellow at Punjab Education Commission (PEC), Govt. of the Punjab, Lahore, Pakistan. He has immeasurable experience of large scale assessment. He is HEC recommended supervisor for doctoral research students. He has worked as national consultant for analyses and interpreting large scale students' assessment data of different sister organizations of National Education Assessment System (NEAS).