An Inquiry on the Roles of Personal Test Item Banking (PTIB) and Table of Specifications (TOS) in the Construction and Utilization of Classroom Tests

Richard R. Jugar
Corresponding Author
Science and Mathematics Education Department
College of Education, University of San Carlos
Cebu City, Philippines

Postal Address:
Room D309 Science and Mathematics Education Department
University of San Carlos, Talamban, Cebu City, Philippines 6000

Contacts:
Phone: +63-32-3443512; +63-32-230-0100 loc. 146; +63933-720-3577
Email: rrjugar@usc.edu.ph

An Inquiry on the Roles of PersonalTest Item Banking (PTIB) and Table of Specifications (TOS) in the Construction of Classroom Tests

Abstract

Classroom assessment, particularly the practice of testing, still exists and continues to be utilized in almost all schools despite the recent advances of alternative assessment schemes. This paper explores the practice of personal test-item banking (PTIB) and the development and utilization of the table of specifications (TOS) by practicing secondary science teachers vis-à-vis the construction of classroom tests. Results indicate that PTIB largely dictates the construction of classroom tests in actual practice without regard to existing principles of classroom test development. As regards to the construction and utilization of TOS, the process is done by working backward contrary to what relevant literatures suggest. The implications on the validity of the developed test items and the accountability of the test results are discussed.

Keywords: *Classroom Test, Test Banks, Table of Specifications, Assessment, Validity, Accountability*

1. Introduction

Educational assessment is the collective practice of making professional judgment and interpretation of the teaching-learning process that encompasses all facets of interaction (Rudner and Schafer, 2002). One of the most dominant forms of assessment used in the context of formal education is classroom testing.Aside from being the most widely used vehicle for assessing (Raagas, 2007), tests can also be viewed as any of a variety of techniques that can capture what an individual knows or understands in response to a specific question (Rudner and Schafer, 2002). As for teacher-made tests, substantial studies have been made concerning its strengths and weaknesses. Marso and Pigge (1992) summarized the findings of over 200 studies concerning teacher-made tests and concluded that the constructed tests depict lack of teachers' knowledge and training, as well as lack of expertise and support. Furthermore, teachers view these tests as supportive of learning but oftentimes contain faults, almost exclusively dwell on the lower cognitive levels, and do not utilize test-improvement mechanisms such as item analysis and test-item blueprints. A subsequent study (Pigge and Marso, 1993) on the attitude of teachers and administrators on teacher-made tests showed that teachers view these tests very positively in light of sound educational assessment practices. Administrators on the other hand, while also expressing relatively positive attitude, only provide limited extent of pre-service and in-service trainings, which will help teachers construct more reliable and valid classroom tests.

While there are merits associated with teacher-made tests (Dereshiwsky, 1993) and attempts are made to create guidelines and practices for its improvement (Williams 1991, Talbot 1994, Torres et al. 2011), more research is needed as to how classroom test-development related practices such as test item banking and table of specifications construction affect the quality of teacher-made tests.

1.1 Test Item Banking

Test item banking (TIB) is defined as the collection or recording of test items that are deemed useful for usage over an extended period of time, as well as for cross and inter-curricular assessments (Freeman et al., 2010).The theoretical underpinnings of the construction of test item banks is usually attributed to Item Response Theory (IRT) for equating different forms of test that can be included in the item bank, although there are existing modifications such as that of Item Banking with Embedded Standards (MacCann and Stanley, 2009). Test item banking or item banking is not only used in the academe but is also utilized in other disciplines. Item banking may also be contextualized to suit a particular need of a specific field. Examples of these discipline-based item banking are the Role Functioning (RF) item bank (Anatchkova, Ware and Bjorner, 2010) and Item Banking cum Computerized Adaptive Test (CAT) among others (Bjorner et al., 2007). In general, test item banking holds the advantages of easier facilitation, rapid result generation and less paper works (Rudner, 1998).

In majority of the extant literature, the context of test item banking is highly institutionalized. In this study, personal test item banking (PTIB) is defined as the individual teachers' practice of collecting both teacher-made and existing tests to be used in classroom testing. PTIB inherently differs from conventional test item banking since the teacher "owns" the bank and is relatively free to choose which item to keep or to discard. Moreover, PTIB is highly personal and

the institution or school does not necessarily monitor nor interfere in the selection of test items for classroom testing as part of the teacher's autonomy of instructional planning (Ozturk, 2012).

1.2 Table of Specifications

Table of Specifications (TOS) or test-item blueprint is a tool in test construction that is used to improve the validity of teacher-made tests by aligning objectives, instruction and assessment. A TOS may be constructed as a two-dimensional table of two types: the first type, which is the most common, utilizes the topics or content covered on the column and the levels of taxonomy on the row. The second type uses the set instructional objectives on the column instead of the topics or content covered (Raagas, 2007). TOS was explained in detail by Fives and DiDonato-Barnes (2013) in terms of *purpose*, that is to improve validity of a teacher's evaluations based on a given assessment; *construction*, that TOS is done *before* instruction or as part of instructional planning; and *utilization*, where the TOS is used to assign the *number* and *level* of test questions. The use of TOS as a tool of ensuring validity was also the focus of the paper by Notar et al. (2004) based on the premise that student achievement, as a requirement of higher education admission, is largely based on the student's grade point average (GPA), which in turn, is heavily influenced by a large measure of teacher-made tests. Due to its use in increasing the quality of teacher-made tests, TOS is deemed an important practice of test development and should be part of instructional planning and design (Kubizsyn and Borich, 2003).

2. Methodology

The study aims to explore and documentthe practices of science teachers who are currently teaching science subjects in selected secondary schools. Specifically, the study aims to inquire on the role of personal test item banking and table of specifications vis-à-vis the construction of classroom tests.In this context, construction of classroom test does not refer to the formulation of the individual test questions but the process of making a type of test which is a collection of a number of questions. Based on the nature of the research problem, the descriptive qualitative research design was employed. Data were collected from identified respondents who are teaching secondary science regardless of the area of specialization. Consequently, the conclusion will be drawn based only on the group participants' characteristics and context.

2.1 Research Participants

The research participants of the study were practicing secondary school science teachers within two city divisions in Central Visayas, Philippines. These secondary school teachers came from both public and private high schools, which generally follow the science curriculum as prescribed by the Department of Education. There were 72 teacher-respondents in this study. Thirty-two (32) of the respondents teach in public while 40 are employed in private high schools. A total of 16 high schools, 8 public and 8 private, participated in this study. In terms of teaching experience, there were 22 teacher-respondents with 0-2 years of teaching experience, 27 teacher-respondents with 3-6 years of teaching experience, 11 teacher-respondents with 7-10 years of teaching experience and 12 teacher-respondents with over 11 years of teaching experience. With respect to highest educational attainment, there were 32 respondents with graduate units or degree

and 40 with bachelor'sdegree. In terms of eligibility, all 72 teacher-respondents had a license or certification from the Philippine Professional Regulation Commission (PRC) and the Philippine Civil Service Commission (PCSC).

3. Results

3.1 Personal Test Item Banking

Based on the interviews, the 40 baccalaureate and 32 graduate respondents explicitly indicated that they use test item banks to some extent. Further probing indicated that the teachers (72/72) did not make the majority of the test items in their item bank themselves, but were compiled from existing test items. Sources of test item compilations include online test item repositories, national tests, foreign standardized tests, and textbook end-of-chapter tests among others. As it is, all teacher respondents (72/72) also indicated that a few of the items in their existing banks were items constructed by the teachers themselves. When asked for an estimated percentage of their personally constructed tests, the answers ranged from 10 to 15 percent of the total items present in their item banks.

An interesting observation in test item banking which greatly affected the teachers' classification of test item is the absence of deliberate or clear alignment of the teachers' test item and the objective of the class encounter. This practice directly affects the validity of the constructed classroom test since 'what is to be asked need be what was taught'. The following transcripts of an interview of three respondents highlight this disparity in test item preparation using the item banks:

Note: This interview was conducted individually. For each question, the responses of the three respondents are written based on the assigned respondent codes.

Interviewer: So how do you select test items from the item bank?

Teacher B5:I usually choose by topic sir. First I make an outline of my topic then distribute the number of items based on the agreed number of items for the test.

Teacher B19: I list down the topics first then assign the corresponding number of questions for each of the topic.

Teacher G2: I have an item then depending on how thorough the topic is discussed, I usually vary the numbers accordingly and mix the easy and difficult ones as part of the test.

Interviewer: How about the objectives for each topic? Do you use these objectives as part of the reference in choosing the items?

Teachers B5 and B19: No sir.

Teacher G2: Not really sir but I do check the end-of-chapter excercises of the textbook to check if the items that I have chosen are ok.

Notice in the transcript that the practice of test-item selection from an established test bank is topic-oriented – that is, the topics taken in class seem to be the core or key factor in the selection process. Whereas this practice may have its inherent merits as regards to the curricular and instructional validity of the test, the absence of cross-referencing in terms of the objectives of the lesson conducted within the scope of the identified topics poses a distinct misalignment between testing and instruction. As the most important determinant of test quality, validity needs to address all three aspects of the teaching-learning process that include planning, instruction and assessment. While it is supposed to be a cause of concern, most of the teacher-respondents seemed to be unaware of this and viewed their current practice of the test-item selection process as acceptable and adequate.

3.2 Construction and Utilization of TOS

The selection process of test questions from an item bank is closely tied with the utilization of the Table of Specifications. The Table of Specification or TOS is a table depicting the distribution of test items across content (topic) and domain (level of cognitive activity) and may replace the topics with the actual statements depicted in the lesson plan. TOS is a tool in test-item construction and classroom testing useful in improving the content validity of test questions (Raagas, 2007). In this study, 72 out of 72 or 100% of the teacher-respondents indicated that they know how to construct TOS and do use it but only for the major examinations. Further, all teacher-respondents indicated that the structure of TOS that they use in practice is that of content vs. domain. When probed if the teacher-respondents find the TOS useful, all respondents indicated that TOS is indeed useful but it was found out that there are some mechanics in the TOS construction and application that deviated from the supposedly accepted process. Notar et al. (2004) defined the TOS as a blueprint of constructing test questions in classroom testing. Akin to a blueprint, a TOS serves multiple functions such as to: (a) clearly define the focus and scope of a test, (b) ensure correspondence between learning objectives and student content course, and (c) organize the process of test development to best represent the material covered in the teaching-learning process. Likewise, just as a blueprint is done prior to the construction of a building, the TOS should be prepared prior to the construction or selection of test. Some authors even require that TOS be done before the actual teaching-learning process as part of curricular planning (Kubizsyn and Borich, 2003).

Contrary to the correct process of constructing the TOS first before the test items, all respondents (72 out of 72 or 100%) indicated that the pool of questions in their respective item banks was the main consideration in their classroom test construction and not the TOS. The usual practice of the teacher respondents is that based on the pre-determined total number of items for a given exam, the exact number of test items are extracted from the existing test item bank with the content coverage as the sole determinant factor. After the selection of the test items, the teacher constructs a TOS based on these extracted items. Once the TOS is done, the teacher identifies which of the domains/topics lack or exceed of the prescribed number of items and consequently changes the test. This is done without regard to the instructional objectives for the content covered. Further, due to the different interpretation of teachers on the different cognitive levels in the domain, the assignment of the items per topic on the appropriate domain is likewise affected. The TOS is

usually submitted as part of the requirements in the construction of major examinations such as quarterly or periodical tests forbasic education in the Philippines. The TOS is submitted to the immediate supervisor (e.g. academic coordinator) for checking; and at times, these supervisors transfer the teachers' choice of domain based solely on how the question is phrased and again, with no regard to the instructional objective nor of the actual teaching-learning process. This practice of checking or correcting the already incorrect TOS does not really improve the validity of the test in question, which is a fundamental deviation from the real purpose of constructing a TOS. The following transcript from an interviewwith one of the graduate teacher-respondents can best illustrate this disadvantage:

Interviewer: So where do you submit your 'completed' TOS?

Teacher G17: We submit it to our academic coordinator, together with the test questions and the answer key.

Interviewer: I see. So does your academic coordinator check your tests?

Teacher G17: Yes, also the TOS.

Interviewer: So how do your tests and TOS get checked?

Teacher G17: The coordinator usually check for the correctness of structure of the test as well as the desiref format of the department. As for the TOS, the coordinator sometimes transfers the domain for a particular item. For example, I put a certain item in comprehension then the coordinator sees it as more like of an analysis, so she advises me to transfer the item.

Interviewer: Oh ok. So what is the basis for the transfer of the item?

Teacher G17: The question. The level of the question I guess.

The transcript depicted above shows one of the misaligned practices of TOS utilization. While it might be true that the attempted correction of the teacher's immediate supervisor could actually improve the classification of the given test items, the absence of cross referencing between and among the written and identified learning objectives pertinent to the topic as well as that of the actual teaching-learning experience may do more harm than good in this case. To illustrate this point, a teacher may have explicitly defined or stated the differences between static and kinetic friction during the actual class encounter. Come the test, the teacher may have an item wherein the students are asked to compare and contrast static and kinetic friction. While compare and contrast might sound like a high-level cognitive activity, the fact that the teacher had explicitly stated the comparison and contrast of both static and kinetic friction in the classroom would render the test item as a simple recall- that is, an item belonging to the lower levels of the cognitive domain. Since the teacher's immediate supervisor may not necessarily observe the actual practice of teaching the concept, a conflict of test item classification may eventually occur. In effect, the already-unclear interpretation of the teachers themselves of the cognitive levels may further be aggravated by this

practice of 'checking' the correctness of the prepared TOS. Overall, the purpose and function of TOS will be for naught and the process of its construction will simply be another paperwork on the part of the practicing teachers with no academic significance whatsoever.

4. Conclusion

The construction of quality classroom tests has its inherent virtues and challenges. While there are established sound principles of test construction and development (e.g. Van den Berg, 2002;Stiggins, 1997), the practices in actual classroom settings may not be necessarily as sound as expected. The practice of PTIB and usage of TOS for instance may have merits on their own, and in theory, are essential tools in the development and construction of sound and quality classroom tests but may prove otherwise in actual practice. These conflicts in theory and practice may be attributed to what some authors call as the paradox of assessment (Katz, Earl and Olson, 2002). The study has provided a glimpse of the role of teachers' PTIB as a major factor and qualitydeterminant of constructed classroom tests, as well as the modifications of TOS construction and utilization that largely deviate from what literature suggests, which as a consequence, renders the tool inappropriate and ineffective.The findings concerning these practices should be an essential consideration in both pre-service and in-service teaching training and enrichment programs for the overall improvement of the teachers' assessment practices.

References

Anatchkova, M., Ware, J., Bjorner, J. (2010). Assessing the factor structure of a role functioning item bank. Quality Life Research, Vol 16, pp. 95-108 DOI 10.1007/s11136-010-9807-1

Bjorner, J., Chang, C., Thissen, D., Reeve, B. (2007). Developing tailored instruments: item banking and computerized adaptive assessment. Quality Life Research, Vol 16, pp. 95-108 DOI 10.1007/s11136-007-9168-6

Dereshiwsky, M. (1993). When "Do it Yourself" Does it Best: The Power of Teacher-Made Surveys and Tests. Paper presented at the Honors Week Symposium. Flagstaff, AZ. Retrieved October 13, 2011. EBSCO Database.

Fives, H., DiDonato-Barnes, N. Classroom Test Construction: The Power of a Table of Specifications. Practical Assessment, Research and Evaluation 18(3). Retrieved Novemner 2013 from http://pareonline.net/pdf/v18n3.pdf

Freeman, A., Nicholls, A., Ricketts, C., and Coombes, L. (2010).Assessing the Feasibility of aTest Item Bank and Assessment.ClearinghouseMedical Teacher, Vol. 32 Issue 6, p464-466. 3p. 4 Charts. DOI: 10.3109/0142159X.2010.486056.

Katz, S., Earl, L., Olson, D. (2001).The paradox of classroom assessment: A challenge for the 21st century. McGill Journal of Education; Winter 2001; 36, 1; ProQuest Education Journalspg. 13

Kubiszyn, T., & Borich, G. (2003) Educational Testing and Measurement: Classroom Application and Practice. (7th ed.) New York: John Wiley & Sons, Inc.

MacCann, R., Stanley, G. (2009). Item Banking with Embedded Standards. Practical Assessment, Research and Evaluation 14 (17). Retrieved November, 2013 from http://pareonline.net/pdf/v14n17.pdf

Marso, R., Pigge, F. (1992). A summary of Published Research: Classroom Teachers' Knowledge and Skills Related to the Development and Use of Teacher-Made Tests. Paper presented at the annual of the Midwestern Educational Research Association, Chicago, IL. Retrieved October 13, 2011. EBSCO Database.

Notar, C., Zuelke, D., and Wilson, J. (2004).The Table of Specifications: Insuring Accountability in Teacher Made TestsJournal of Instructional Psychology, v31 n2 p115-129 (EJ774063)

Ozturk, I. (2012). Teacher's Role and Autonomy in Instructional Planning: The Case of Secondary School History Teachers with regards to the Preparation and Implementation of Annual Instructional Plans. Educational Sciences: Theory and Practice 12(1) pp. 295-299

Pigge F., Marso, R. (1993). A summary of Published Research: Classroom Teachers' and Educators' Attitudes Toward the Support of Teacher-Made Testing. Paper presented at the annual of the American Educational Research Association, San Francisco, CA. Retrieved October 13, 2011. EBSCO Database.

Raagas, E. (2007). Assessment and Evaluation of Student Learning: Concepts and Applications. 2nd Ed. DATStat Analysis Center, Cagayan de Oro, Philippines

Rudner, Lawrence (1998). Item banking. Practical Assessment,Research &Evaluation, 6(4). Retrieved November 2013 fromhttp://PAREonline.net/getvn.asp?v=6&n=4.

Rudner, L., Schafer, W. (2002). What Teachers Need to Know About Assessment.National Education Association, Washington D.C. Educational Resource Information Center (ERIC). Retrieved October 13, 2011. EBSCO Database.

Stiggins, R. (1997). Student-centered classroom assessment. 2nd Ed. Upper Saddle River, NJ: Prentice-Hall Inc.

Talbot, G. (1994). Revitalizing Teacher-Made Tests: Quality Control Procedures. Educational Resource Information Center (ERIC). Retrieved October 13, 2011. EBSCO Database.

Torres, C., Lopes, A., Lurdes, B., Jose, A. (2011). Improving Multiple-Choice Test Questions. US-China Education Review B 1 (2011) 1-11. Retrieved, October 13, 2011. EBSCO Database.

Van Den Berg, E. (2002). Assessment in Science.Book (Unpublished).Science and Mathematics Education Department, University of San Carlos. Cebu City, Philippines.

Williams, J. (1991). Writing Quality Teacher-Made Tests: A Handbook for Teachers. Educational Resource Information Center (ERIC). Retrieved October 13, 2011. EBSCO Database.