SOUTHEAST ASIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

Does this OBE Count?: Test Development and Item Analysis in High School Dressmaking Course

Jhoysa C. Camba[1], Rosanta V. Alibin[2], Jun S. Camara[3], Annie M. Calicdan[4], Amelita M. De Vera[5], Rosario DL Valencerina[6]

[1]Bachelor of Technical Vocational Teacher Education, College of Education
[2]Faculty, BTVTEd, Technical Education and Specialization Course Department
[3]Faculty, Assessment in Learning 1 & Campus Research Coordinator
[4]Faculty, BTVTEd, Technical Education and Specialization Course Department
[5]Department Chair, Technical Education and Specialization Course Department
[6]Faculty, Professional Education Department, College of Education
[2-6]Faculty, College of Education, Pangasinan State University, LINGAYEN Campus
Correspondence: jhoysacamba@gmail.com

Abstract

If you know how to do it, then do it. This developmental study, generally, explored on the competency of BTVTEd pre-service teachers in test development and item analysis. On a more particular sense, their classroom-based output (i.e. a Summative Test in Dressmaking) was developed, converted as a google form, and administered to TVL students in a national high school in Pangasinan, Philippines. The summative test went series of revisions and validation that enabled both the students and faculty experts to explore how OBE works in assessment course. The test included fifty items under Dressmaking and was administered digitally. Item analysis of the index of difficulty used Henning's (1987) and index of discrimination used Ebel's (1979) scale interpretation. Findings showed that 18% of the items were 'retained' while majority of the items were ordered for 'revision'. These findings could provide a baseline information, if not a fit comparison, on test-taking between face-to-face and online test administration. Further, the retained items could be evaluated to shed light on which topical competencies are answered correctly with varying difficulty against increasing discriminatory index, to enable a closer look on the context where the students are coming from. Distractor analysis is recommended as well as integrating a gender lens to the items which were considered for 'revision' to shed light on whether gender stereotyping in 'dressmaking' influences the results.

Keywords: *Assessment, Dressmaking, Technical Education, TOS, Test, Senior High School*

INTRODUCTION

Nearly everyday, classroom teachers write assessment tools, i.e. formative and summative, to measure pre- and post-learning of learners. What makes this a dauting task for teachers is the preparation of test questions every year in order to avoid duplication of test questions in the previous year and in the next years to come even if learning competencies do not actually change (Camara, 2021). In a study by Plake in 1993, competencies of teachers based on a national survey *(n=555)* showed that teachers had their best performance in assessment (i.e. administering, scoring, and interpreting test results) but had the worst performance (i.e. communicating test results) as well in assessment. Very significantly at this present time, Chan & Luk (2021) in a nationwide study *(n=2150)* found that both teachers and

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

students now have common views towards holistic competency assessment and grading preferences, whereby students prefer holistic grading but go against making quantitative scoring of holistic competencies, and they preferred having a qualitative record of their holistic competency without grades attached. In the case of the Philippines, these issues on holistic assessment and non-score grading are not yet established, and the introduction of these policies have yet to be considered, against very important norms including society, culture, internationalization, and many more.

In the context of the Philippines, the teacher as the most important instructional material inside the classroom remains the authority in the learning of his or her students, much more with issuing student scores and grades based on pre-determined sets of test, activities, performance, and other assessment modalities the teacher deems fit. With this set-up, the teacher is given the responsibility of developing both instructional materials and assessment modalities that are tailor-fit to his or her students. Relative to this is the competency that the teacher must possess in constructing test questions, administering the Test, analyzing scores to improve learning, and communicating results to both students and parents. This depiction of a Filipino classroom teacher in terms of assessment provides a clear boundary that teachers do need to develop tests, among others, for classroom use. In the construction of the Test, though, he or she has to target only the DepEd- or government-mandated learning competencies which students need to master for each year level in a span of one year, in four quarters.

One of the major problems in classroom assessment, particularly with today's blended learning set-up of instruction is the need to measure the extent to which skill competencies are measured even when the execution of these skills are not practiced inside the classroom, especially in skill subjects like Dressmaking. The very few studies literature in the Philippines on available summative test for use by Dressmaking students under the TVL track inspired the researchers to develop and validate a Summative Test based on pre-determined topical competencies. Further, this study could provide a working strategy in the institutionalization of practical test development and validation among assessment courses in higher education institutions.

METHODOLOGY

Developing the Simplified TOS for METD

This action research started with identifying the most essential topics in the DepEd's Dressmaking G7 and G8 Course Guide by the students and faculty-experts specialized in Teaching Garments, Fashion and Design (GFD) from the Technical Education and Specialization Course of Pangasinan State University, Lingayen Campus, through an informal interview. Careful validation by faculty experts resulted to eleven (11) Most Essential Topics in Dressmaking (METD).

The rhetoric of framing questions using the Bloom's Taxonomy was not yet considered in this course-based action research because this study was initially initiated to explore on item analysis processes and not on ascertaining how much the prospect examinees have achieved per learning competency or topic. The List of Topical Competency was assigned a specific number of items to be developed based on the consultation of the students with faculty-experts.

Table 1 displays the METD in HS Dressmaking and the number of assigned items for each METD. The total number of items in fifty (50) and are distributed by 8-10-12 scheme implying that each METD is given proportional weights in the Summative Test (ST). The selection of the METD include considerations on Tools, Processes, and Risks in Dressmaking.

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

Tools and Processes are essential questions while risk is included because of the nature of instruction (modular/virtual) brought about by the pandemic.

Table 1: The METD in HS Dressmaking

No	METD	NoI	%
1	Measuring Tools in dressmaking and Different types of sewing machine in dressmaking	6	12
2	Different types of sewing machine in dressmaking	4	8
3	Parts and function of the domestic sewing machines	5	10
4	Taking accurate body measurement in Dressmaking	4	8
5	Calculating simple calculations using metric conversion chart	5	10
6	Principles of design in dressmaking	4	8
7	Basic Hand stitches in dressmaking	4	8
8	Proper handling and cleaning of the Lockstitch sewing machine	4	8
9	Procedures in lower and upper threading of Lockstitch sewing machine	4	8
10	Hazards and risks in the workplace in dressmaking	6	12
11	Causes of hazard and risk that may encounter in workplace in Dressmaking	4	8
	Total	**50**	**100**

Developing the draft of achievement test
Developing the Test Questions for METD

The researchers developed the first draft of the summative test by finding books, eBooks, and other instructional materials on the METD that related to DepEd curricula and ASEAN Integration questions (Calicdan, 2017). Since most of the METDs are content-wise, the researchers explored on converting written paragraphs relative to an METD into appropriate multiple-choice questions. Plausible alternatives were chosen, carefully selected to increase their homogeneity to reduce the risk of guessing among the examinees. Appropriate assessment writing principles were carefully followed. Faculty-experts validated the content and correctness of the answers to each test question as well as the general reliability of the books, eBooks and other IMs consulted. The format was carefully followed in the entire test, which includes the following: (1) the stem should be written in an interrogative form, (2) the options should be homogeneous, arranged appropriately, and are good distractors; and (3) the arrangement of items in the google form could be jumbled so as to make sure that cheating is addressed appropriately.

Reading Score of the Draft Achievement Test

Readability test is also an important step in framing test questions in order to make sure that language or vocabulary words do not become a barrier (Camara, 2021) to understand the test questions, and making sure that the words used in the test questions are within the reading level of the target users, i.e. the Senior HS Grade 11 Students. Based on the Flesh-Kincaid Readability Test, the reading level of the final draft of test was found to be for grade 11. Further, the draft was subjected to similarity check in a Turnitin.com software and yielded a <15% similarity check and making sure that the test questions were developed along the way without reference to questions in dressmaking already available in the world wide web or any online repositories. The Draft Test went series of revisions and content validation.

Test Administration via Google Form

The ST was in the form of a google form in order to reach the target examines. The study locale was in Labrador National High School (LNHS), Labrador, Pangasinan. The test administration was granted permission to

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

conduct by the school officials and was participated in by 80% of the TVL Dressmaking students enrolled in the LNHS. The choice of study locale is under the principle of convenient (i.e. reachable by the lead researcher) and purpose (i.e. the target users are available in the study local) sampling. The restriction of IATF guidelines prohibits non-medical related concerns for a face-to-face setup which left the researchers no choice but to conduct it via google forms. The lack of viable guidelines for a DepEd-SUC collaboration on student activities including the request of test administration will not permit a more massive test administration, and the use of one school is the most viable route at the time of data-collection.

Test Item Analysis

Results of the STs by the examinees were tabulated and tallied via google sheets and were statistically analyzed using the suggestion of Henning (1987) for the index of difficulty and the suggestion of Ebel (1979) for the index of discrimination. According to Henning's, an item is easy if it has a difficulty index of ≥ 0.67, of medium difficulty if the difficulty index is 0.34 – 0.66, and high difficulty if it has a difficulty index of ≤ 0.33. Further, according to Ebel, an item is a very good item if it has a discrimination index of 0. 40 – 1.00, a good item if the index of discrimination is within 0.30 – 0.39, it is a marginal item if the index of discrimination is within 0.20 – 0.29, and a poor item if the index of discrimination is below 0.19.

RESULTS

Report of Analysis

Table 2 shows the indices of difficulty and discrimination of the METD Test Items. A Decision for each MEDT is also given based on these indices.

Table 2: Indices of Difficulty and Discrimination for each METD

No	Diff. in.	Disc. in.	DECISION
1	56.25[m]	-0.13[p]	Revise
2	21.43[h]	0.25[m]	RETAIN
3	28.57[m]	0.00[p]	Revise
4	28.57[m]	0.00[p]	Revise
5	28.57[m]	0.00[p]	Revise
6	28.57[m]	0.00[p]	Revise
7	32.14[h]	-0.13[p]	Revise
8	14.29[h]	0.00[p]	Revise
9	17.86[h]	0.13[p]	Revise
10	21.43[h]	0.25[m]	RETAIN
11	21.43[h]	-0.25[p]	Revise
12	21.43[h]	0.00[p]	Revise
13	10.71[h]	-0.13[p]	Revise
14	17.86[h]	-0.13[p]	Revise
15	21.43[h]	-0.50[p]	Revise
16	10.34[h]	0.13[p]	Revise
17	10.34[h]	-0.13[p]	Revise
18	6.90[h]	0.25[m]	RETAIN
19	20.69[h]	0.00[p]	Revise
20	17.24[h]	-0.13p	Revise
21	10.34[h]	0.38[g]	RETAIN
22	13.79[h]	-0.25[p]	Revise
23	11.90[h]	-0.13[p]	Revise
24	11.90[h]	0.13[p]	Revise

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

25	9.52[h]	0.00[p]	Revise
26	14.29[h]	0.00[p]	Revise
27	11.90[h]	0.13[p]	Revise
28	16.67[h]	0.38[g]	RETAIN
29	11.90[h]	0.13[p]	Revise
30	16.67[h]	0.38[g]	Revise
31	16.67[h]	0.38[g]	Revise
32	11.90[h]	0.63[vg]	Revise
33	4.76[h]	0.00[p]	Revise
34	11.90[h]	0.38[g]	Revise
35	7.14[h]	0.38[g]	Revise
36	17.86[h]	0.13[p]	Revise
37	25.00[h]	-0.13[p]	Revise
38	10.71[h]	-0.13[p]	Revise
39	7.14[h]	0.13[p]	Revise
40	9.52[h]	0.00[p]	Revise
41	21.43[h]	0.25[m]	RETAIN
42	25.00[h]	0.13[p]	Revise
43	10.71[h]	-0.13[p]	Revise
44	7.14[h]	0.00[p]	Revise
45	7.14[h]	0.00[p]	Revise
46	9.52[h]	0.25[m]	RETAIN
47	16.67[h]	0.13[p]	Revise
48	11.90[h]	0.13[p]	Revise
49	11.90[h]	0.38[g]	RETAIN
50	9.52[h]	0.25[m]	RETAIN

Legend:

diff *e*Easy, *m*Medium Difficulty, *h*High

disc *p*Poor Item, *g*Good, *m*Marginal, *vg*Very Good

DISCUSSION

Generally, findings showed that 18% of the test items were decided to be retained in the ST, while majority of the test items, 82%), were all considered for revision. It should be noted that the authority to revise, retain, and reject a test item emanates from the classroom teacher in a criterion-reference test because of the principle that they are the ones exposed to the nature of their learners and they could rightfully say so when an item possible for rejection could still be considered for revision. Camara (2021) emphasized that the role of teachers in the assessment of student learning is indispensable to nation-building because they make sure that the target competencies are learned at least at a level desirable by the society as contained in its mandated school curricula. Further, he noted that the role of the teacher to cause the validation and/or themselves as validator reflects an image of high reciprocity between and among them, professionally and academically.

Susie (2017) pointed out that in any case, all indices should be considered together before making decisions or revisions. One important thing to always keep in mind is that decisions about item revision should be based on the extent to which item performance matches your intent for the item and your intent for the overall exam. Both of these suggest that while the teacher has all the pieces of evidence to improve learning, consideration of several other things would be considered. The number one to be considered is the nature of the both the learner and the environment which, when taken together, put assessment in a higher playing field. In the context of the findings, the prevailing skill-focused scheme between and among the dressmaking students (Camara, 2018; Paborada & Valencerina, 2019) may have caused the high

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

difficulty found among the items, since it is a general notion that skill subjects normally focus on skill performance rather than on rote learning. Further, the prevailing policy of Philippine's Department of Education of non-periodic test since the beginning of the COVID-19 pandemic may relaxed the intellectual exercise among the students that what they were taking was an actual test and not a mere summative test taken from a unit, chapter or a book. After all, Mehrens & Lehmann (1973) cautioned that item analysis data are tentative. Such data are influenced by the type and number of students being tested, instructional procedures employed, and chance errors.

Furthermore, this study aimed to test the principles of test construction and test item analysis among conveniently sampled participants with a purpose, and the ST administered may undergo further revision to achieve a level digitally feasible for summative test to target users (i.e. dressmaking students). The validity of the test in the eyes of the faculty-experts may not have been confirmed by the indices of difficulty and discrimination as shown in Table 2 but it was able to ascertain some items (18%) of them which could be used as pattern to fully understand the intellectual behavior of the students. The intellectual behavior of the students in this sense simply refers to the analysis of the topical competencies which are normally found difficult by the examinees but were able to discriminate them into 2 categories, first from those who really know the tested material and from those who did not and may have probably just guessed about in choosing an answer.

CONCLUSIONS

On a general note, this study aimed to develop and validate a Summative Test in Dressmaking for use by Senior High School. On a deeper note, this study has attempted to explore on the skills of BTVTEd Pre-Service Teachers in Assessment in Learning 1 particularly in Test Construction, Test Validation, and Test Item Analysis, by putting into practice and outside the classroom the very foundational assessment principles. Not only will this enhance the would-be practice teachers as Calicdan (2017) emphasized, but all the more capacitate future THE or TLE teachers as noted by Paborada & Valencerina (2019).

The findings of the study that revealed the difficulty and discrimination indices of the entire summative test package is a manifestation that what could be valid in the eyes of faculty-experts may not work in the eyes of the students in the 'real world' and therefore continuous curriculum assessment is necessary as Camara (2020) observed that no curriculum exists in isolation, i.e. each curriculum is designed to be an input to another output and that the interplay of curricula in the trifocal system of education in the Philippines only becomes ideal when alignment is checked at crucial checkpoints, one during the implementation of K to 12, and today during the implementation of modular distance education or other forms alternative to holding classes face-to-face.

Further, in an exploratory sense, the success of the study rests on the timely completion of the student-researchers (i.e. Team of writers) with the series of draft of test questions, simplified TOS, and series of tabulation, analysis and interpretation of 'numbers'. This implies that the use of this strategy in the teaching of Assessment in Learning 1 is feasible in a semestral scheme.

Furthermore, the Summative Test could undergo Phase 2. In Phase 2, the 18% of the items that were ordered RETAINED shall be further analyzed in terms of distractor analysis. The rest of the items that were ordered

Volume 6, Issue 1, 2021
P-ISSN: 2672-2984
E-ISSN: 2672-2992
www.sajst.org

REVISED shall undergo further content-validation from teachers in DepEd schools which offer Dressmaking courses to further contextualize the items. Finally, for Phase 2, it could include the variable on gender perception on dressmaking and note whether female students would tend to have higher scores compared with their male classmates to at least determine if gender stereotyping is a moderating variable in dressmaking summative test. After all, this OBE counts as a way to explore its integration among assessment courses in higher education institutions.

LIMITATIONS

The use of the achievement test may find its limitations in terms of general applicability because learning competencies targeted were randomly selected based on DepEd curriculum guide and the relatively low number of examinees, who unfortunately were the 'everyone' there were already from the TVL strand at the time of data-collection. This action research aimed to test theory in Assessment in Learning 1 into classroom-based practice by pre-service BTVTEd Pre-service teachers. Further, this write up is an output of the 4AA Model developed by Dr. JS Camara to reduce academic writing requirements by focusing on a task-based approach to manuscript writing.

ACKNOWLEDGEMENTS

The authors would like to thank Labrador National High School of Labrador, Pangasinan, Philippines for accommodating the request of the researcher to administer the Final Draft of the Test Questions among their TVL students, and for their suggestions and recommendations to the Table of Specifications.

REFERENCES

Calicdan, S. (2017). Implementation of the Practice Teaching Program of the Pangasinan State University. *Southeast Asian Journal of Science and Technology, 2(1), 12-16*

Camara, JS. (2021). Development and Validation of a Criterion-Referenced Test in Qualitative Research for 11th graders. *Available at SSRN 3865856*

Camara, J. S. (2020). Philippine Biology Education for a Curricular Innovation towards Industrial Revolution 4.0: A Mixed Method. *Asian Journal of Multidisciplinary Studies*, *3*(1), 41-51

Camara, J. S. (2018). Correlates of Self-efficacy, Learning Style and Aptitude Across Strand of Senior High School Students in San Jacinto National High School. *Asian Journal of Multidisciplinary Studies*, *1*(2), 15-24

Chan, C. K., & Luk, L. Y. (2021). A four-dimensional framework for teacher assessment literacy in holistic competencies. *Assessment & Evaluation in Higher Education*, 1-15

Mehrens, WA and I. J. Lehmann. (1973). Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, 333-334

Paborada, M. N., & Valencerina, R. D. (2019). Competencies in Home Economics of Grade 10 Students in Mangatarem District. *PSU Multidisciplinary Research Journal*, *2*(1), 56-6

Plake, B. S. (1993). Teacher Assessment Literacy: Teachers' Competencies in the Educational Assessment of Students. *Mid-Western Educational Researcher*, *6*(1), 21-27

Suskie, L. (2017). "Making Multiple Choice Tests More Effective." Schreyer Institute for Teaching Excellence. The Pennsylvania State University

Understanding Item Analyses. (2018). Office of Educational Assessment. University of Washington. Retrieved from http://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/