

Checklist versus Rating Scale in Psychomotor Assessment: Achieving Objectivity

¹IBEZIM, N. E. and ²IGWE, N

¹ Department of Computer Education, University of Nigeria, Nsukka

² Department of Industrial Technical Education, University of Nigeria, Nsukka

ABSTRACT: Checklist and rating scale are instruments used in assessing students in psychomotor domain. One of the challenges in psychomotor domain is to achieve objectivity; so that any rater would give a score which another rater can also arrived at, thereby achieving reliability and validity which are vital in any test. This paper x-rays checklist and rating scale to ascertain which of them would be better in achieving objectivity. Also the paper suggests ways to achieve objectivity using such instrument.

I. Introduction

Checklist and rating scale are instruments used to assess psychomotor domain. A checklist is both a reporting tool and an assessment tool. As an assessment tool, one can easily adapt a checklist to a grading scale by assigning points or weights to the criteria on the checklist (Karges-Bone, 2000). A sound evaluation checklist clarifies the criteria that at least should be considered when evaluating something in a particular area; aids the evaluator not to forget important criteria; and enhances the assessment's objectivity, credibility, and reproducibility (Stufflebeam, 2000).

On the other hand, Encyclopaedia Britannica (2016) defines the rating scales as that which present users with an item and ask them to select from a number of choices. The rating scale is similar in some respects to a multiple choice test, but its options represent degrees of a particular characteristic. Razzi (2011) stated that a rating scale is a commonly used scale system for performance appraisals. The scale typically features a Likert scale from 1-3, 1-5, and so on. Performance appraisals, just as in cognitive appraisals should be objective.

An objective test item is defined as one for which the scoring rules are so exhaustive and specific that they do not allow scorers to make subjective inferences or judgments; thereby, any scorer that marks an item following the rules will assign the same test score. Objective test's high reliability and predictive and content validity led to the gradual replacement of the essay test (Murayama, 2009). Slack (2014) added that a test that is objective measures without reference to outside influences. Irrelevant, unrelated factors do not influence the test results if a test is objective. FitzGerald (1999) also stated that in testing, objectivity means simply universal agreement. A test is objectively scored if everyone scoring it arrives at the same score. Tests like essay examinations, though will obviously fall short of this ideal; proper training of scorers followed by proper monitoring of scoring can produce essay examinations which approach total objectivity. Nevertheless, they will not approach it closely and the interpretation of essay test results, or of the results of any test which is less than completely objective, should include consideration of statistical analysis of its objectivity. Irrespective of the level of essay examination towards objectivity, it cannot be as objective as a conventional objective test. Conventional objective test is of all selected response format, which are true/false, multiple choice, and matching test items.

In explaining objective test formats, Murayama (2009) affirmed that a variety of different types of objective test formats can be classified into two categories: a selected response format, in which examinees select the response from a given number of alternatives, including true/false, multiple choice, and matching test items; and a constructed response format, in which examinees are required to produce an entire response, including short answer test items. The true/false test is the simplest form of selected response formats. True/false tests are those that ask examinees to select one of the two choices given as possible responses to a test question. The choice is between true and false, yes and no, right and wrong, and so on. A major advantage of the true/false test is its efficiency as it yields many independent responses per unit of testing time. Therefore, teachers can cover course material comprehensively in a single test. However, one apparent limitation of the true/false test is its susceptibility to guessing. It should be noted, however, that test givers can attenuate the effects of guessing by increasing the number of items in a test. In addition, some guessing might reflect partial knowledge, which would provide a valid indication of achievement.

Another selected response format type is the multiple-choice test, which has long been the most widely used among the objective test formats. Multiple-choice test items require the examinee to select one or more responses from a set of options (in most cases, 3–7). The correct alternative in each item is called the answer (or the key), and the remaining alternatives are called distracters. Examinees have less chance of guessing the correct answer to a multiple-choice test question compared to a true/false test question. In addition, the distracter

an examinee selects may provide useful diagnostic information. Related to the multiple-choice test is the matching test, which consists of a list of premises, a list of responses, and directions for matching the two. Examinees must match each premise with one of the responses on the basis of the criteria described in the directions. A major strength of the matching test is that it is space-saving and, therefore, can be used to assess several important learning targets at once.

A typical example of a constructed-response format is the short-answer test, which asks examinees to supply a word, phrase, or number that answers a question or completes a sentence. Sometimes it is called a completion or fill-in-the-blank test. Although what a short-answer test item can assess is generally more limited to factual information, it does not require the development of plausible distracters. Moreover, short-answer items are much less susceptible to guessing than selected-response format items. Conventional objective test are more popular in cognitive domain, but it can also be used in affective and psychomotor domain.

Checklist, which is popular test instrument in psychomotor domain, can be customized into a conventional objective test by reducing the options the examiners have to only two; basically right and wrong options. The basic difference between cognitive true/false and that of psychomotor is that the examinee responds to the question in cognitive domain, while the examiner (or his/her delegate) responds to the checklist in psychomotor domain. Also in psychomotor domain, true/false guessing is attenuated. This is possible because the examiner should be aware of all the correct responses expected of the examinee, unlike the cognitive domain where the examinee is being examined. This makes a checklist psychomotor testing more objective, as it reduces its shortcomings over the cognitive test.

Psychomotor testing (performance testing/assessment) has some shortcomings over cognitive testing, which makes the latter more popular. Psychomotor test is a test determining the efficiency of use of cognitive and motor skills (Psychology Dictionary, 2016). Okoro (2002) identified some of the shortcomings of performance testing. First, performance tests are costly to administer because of the tools, equipment and materials needed. Secondly, performance tests require much of the teacher's time in setting up, administering and grading the test. Next, performance test may test only a small sample of the skills possessed by students. If most of the required skills are not tested due to limitations of equipment or time, this could lead to low level of validity and reliability of performance tests. Again, bias of the rater, especially in the case of process measurement, can lead to unreliability of performance tests. FitzGerald (1999) added that the problem with performance assessment is that it is difficult to devise dependable tests of this type. Research on educational performance assessments has shown that tests of similar topics often produce dissimilar results. These situations are so because rating scale is used to test most psychomotor test; and rating scale is subjective rather than objective. There is therefore the need to reach objectivity in psychomotor testing. In order to achieve objectivity in psychomotor testing, checklist should be used in assessing the students' psychomotor performance.

The problems identified above are cost in money and time, lack of content validity and subjectivity. Checklist which has two options can be used to solve the problems. This is because a checklist of two options qualifies to be an objective test -- true/false objective type. For any objective test to be convincing, it should possess objectivity, reliability, validity and economy. A convincing objective test eliminates the identified problems of psychomotor testing. Explaining the four criteria Nolte (n.d) stated that an objective test measures a variable independent of the people conducting the test and the circumstances of the test. A reliable test accurately measures a quality and is repeatable. A valid test measures a specified ability, which includes content validity. Content validity is ascertaining whether what the test set out to test is what it is testing. A test is economical if its overall costs are manageable for the programme. Those costs could be money but also include time involved, necessary equipment, and personnel.

An economical criterion is particularly important when measuring a teacher made psychomotor test in a classroom. The format of end of programme assessment drives the format of teacher made test. Since most psychomotor end of programme assessment is in subjective format, the teachers also set their questions in that line. This makes the teachers not to attain content validity in psychomotor testing. Content validity is particularly important in any curriculum. In order to attain content validity, FitzGerald (1999) stated that Performance assessments are often useful when you can make a lot of performance assessments. Since it is costly to assess psychomotor skills, there is the need to break the entire psychomotor curriculum into single periods. The present practice is integrating psychomotor lesson into cognitive lesson. This practice should be phased out in order to accommodate the uniqueness of psychomotor lesson. It should be observed that there is different classification of cognitive domain from psychomotor domain. The differences should also reflect in the teaching. The principle to teach from known to unknown in cognitive domain, using Bloom's cognitive Taxonomy (1971), would be to teach for knowledge, then to comprehension, application to evaluation. This is the basic steps to take while drafting a lesson plan. In Dave's (1975) classification of psychomotor domain, applying the same principle would be to encourage the student to imitate, and then manipulate to naturalize. It therefore demands different approaches in imparting the principles. At any level in the psychomotor assessment checklist should be used. Using checklist would same time for the teacher in order to attain content validity.

Introducing only checklist in assessing the students in psychomotor domain would also change the idea of norm reference as is presently used in classroom. Norm referencing is basically to classify students according to who is better academically, and mostly a cognitive test serves the purpose; whereas the idea of performance assessment is simply that the best way to assess someone's mastery of a skill is to have them perform the skill. For example, people are not allowed to drive on the public highways until they have passed the driver's test (FitzGerald, 1999). Slack (2014) also opined that a driving test that only measures knowledge of traffic laws is not a valid measure of driving ability since the written test alone does not adequately assess all skills required to be a successful driver. A driver's licence is not issued to someone simply because he performs better than others in the batch of mates who learnt the driving with him; rather it is given to only individual who have satisfied the entire basic requirement stated in the driving checklist. This means that their performances were compared against a criterion rather than a norm-referenced standard.

This idea of using norm-reference standard should also be brought into the classroom. A two option checklist of the requirement of all the skills, in its progressing level of difficulty, in any psychomotor module should be outline before the commencement of any programme. The students should be assessed based on that. Each student should be encouraged to progress at his own pace. The skill acquired at the end of a session could even be used in ascertaining a better student, if the teacher so desire. This could reduce the unnecessary pressure on the classroom teacher to assess for content validity in psychomotor assessment which has been elusive using rating scale.

An objective psychomotor testing would also increase its popularity in experimental research. Most classroom experimental research is done as quasi experimental research. According to Donald, Lucy, Asghar and Chris (2006) quasi-experimental design are similar to randomized experimental designs in that they involve manipulation of an independent variable but differ in that subjects are not randomly assigned. Because the quasi-experimental design does not provide full control, it is extremely important that the researcher be aware of the threats to both internal and external validity and considers these factors in the interpretation. Gall, Gall and Borg (2007) added that this design is important where it is necessary not to disrupt the academic programmes of the school involved in the study. Intact classes would be used in this situation. The researcher also ensures that he/she controls the threats in order to arrive at internal validity.

One of the threats to control is teacher variable. Teachers possess different standards which include knowledge of the content and methodology. These different standards make them assign different scores for the same assessment when subjective test is used in assessing the student. There is therefore the need to use only objective checklist in assessing the student's psychomotor domain. The classroom teachers are used in quasi-experimental research (in order to control subject effect). Subject effect is a situation when the subject (in this case, the students) is aware that they are in experimental research. In order to avoid experimental bias, the researcher would involve the services of the same classroom teachers in handling experimental and control groups.

Objective psychomotor assessment would also make it possible to compare computer assisted instruction for psychomotor domain with non-computer psychomotor instruction. Lieve, Didier, Lise, Filip, Wouter and Bernard (2015) observed that computerized task is an objective and reliable method to assess fine motor activity, and is generally considered to be a rater-independent and more objective measurement method than the more subjective rating scales. Moreover, rating scales can be rater dependent whereas the currently used assessment method (the use of computer and digital tools, such as the simulation) does not depend on the rater. Psychomotor variables are digitally programmed, calibrated and recorded in digit tools. Therefore, when a two option checklist is used to assess the psychomotor domain it becomes on a par with the computerized instruments in terms of comparison; thereby making an experimental research of the two teaching techniques feasible and reliable.

Reference

- [1] Bloom, B. S. (1971). *Taxonomy of educational objectives*. New York: Longman Publishers. Retrieved from <http://teaching.unc.edu/learning-resources/articles-books/best-practice/goals-objectives/b..>
- [2] Dave, R. H. (1975). *Developing and writing behavioural objectives*. Retrieved from <http://www.ucd.ie/t4cms/ucdtla0034.pdf>
- [3] Donald, A., Lucy, C. J., Asghar, R., & Chris, S. (2006). *Introduction to research in education*. Canada: Thomson Wadsworth
- [4] Encyclopaedia Britannica (2016). *Personality assessment: Rating scale*. Retrieved from <http://www.britannica.com/topic/rating-scale>
- [5] FitzGerald J. (1999). *Testing the tests*. Retrieved from <http://www.actualanalysis.com/tests.htm>
- [6] Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction*. Boston: Pearson Educational Inc.
- [7] Karges-Bone, L. (2000). *A checklist for everything! : Simple assessment tools for student projects, grants and parent communication*. Retrieved from <https://www.ocps.net/cs/ese/support/curriculum/Documents/A%20Checklist%20for%20Everything%20Book.pdf>
- [8] Lieve L. B., Didier S., Lise D., Filip B., Wouter H. & Bernard S. (2015). *Psychomotor retardation in elderly untreated depressed patients*. *Frontiers in psychiatry*. doi: 10.3389/fpsy.2014.00196
- [9] Murayama, K. (2009). *Objective test formats*. Retrieved from <http://www.education.com/reference/article/objective-test-items/#B>

- [10] Nolte V.(2nd ed.) The five factors to consider when selecting test procedures. Retrieved from <http://www.humankinetics.com/excerpts/excerpts/the-five-factors-to-consider-when-selecting-test-procedures-excerpt>
- [11] Okoro, O. M. (2002). Measurement and evaluation in education. Anambra pacific Publishers Ltd
- [12] Psychology Dictionary (2016). What is psychomotor test? Retrieved from <http://psychologydictionary.org/psychomotor-test/>
- [13] Razzi, K., (2011) Graphic rating scale. Retrieved from <http://blog.tnemployeeinsights.com/graphic-rating-scale/>
- [14] Slack, K. J. (2014). Targeting objectivity, reliability, and validity. Retrieved from <http://www.thewisdomthatworks.com/targeting-objectivity-reliability-validity/>
- [15] Stufflebeam D. L. (2000). Guidelines for developing evaluation checklists: The checklists development checklist (cdc) Retrieved from https://www.wmich.edu/sites/default/files/attachments/u350/2014/guidelines_cdc.pdf