



DEVELOPMENT AND VALIDATION OF CLASSROOM OBSERVATION TOOL FOR FACULTY TEACHING UNDERGRADUATE COURSES

Dr. Gregg O. Siat

Biliran Province State University, Naval, Biliran, Philippines

greggsiat@gmail.com

DOI: 10.47760/cognizance.2023.v03i11.019

Abstract: The purpose of this study is to determine the validity and reliability of the test items developed for classroom observation for faculty teaching undergraduate courses. The study utilized semantic validation, content validation and reliability testing to ensure relevance and clarity of the items. The respondents of the study were the program chairperson, school dean, university psychometrician, and Faculty Federation President. The finding of this research indicated that majority of the items developed for classroom observation were relevant and clear that will give justice and compensate the actual teaching-learning performance of the teachers.

Keywords: Teaching-learning process; classroom management; teachers' performance

Introduction

Classroom observation is a method of directly observing teaching practice as it unfolds in real time, with the observer or analyst taking notes and/or coding instructional behaviors in the classroom or from videoed lessons. Though widely used across the educational spectrum, the technique is far more common—and the methodological sophistication more pronounced—in K–12 schools, where protocols such as the Classroom Assessment Scoring System (Pianta, La Paro, & Hamre, 2007) and the Framework for Teaching (Danielson, 2007) are used for teacher evaluation.

Classroom observation has many valid and important educational purposes. It is widely used as description of instructional practices; investigation of instructional inequities for different groups of students; and improvement of teachers' classroom instruction based on feedback from individual classroom or school profiles.

There are three main reasons schools observe their teachers - and why they'll likely continue to do so. Classroom observations, when done right can improve a teacher's ability to teach and, as a result, improves student outcomes as well; allow schools to investigate and/or identify potential bias in how different groups of students are taught and treated. Of particular interest are the differentiating factors

of gender, socio-economic standing, religion and race; and provide researchers with current information on both educational practices and instructional problems.

According to Williams, classroom observations should be “developmental rather than judgmental” in the sense that they offer opportunities for teachers to improve their awareness, abilities to interact and evaluate their own teaching behaviors’. There is no doubt that the classroom is a place where many processes of learning and teaching occur. In this respect, it is extremely relevant to consider what to observe and how to observe it (Wajnryb, 1992). What the teacher does and what really happens in the foreign language classroom are what is usually observed. However, apart from the teacher and his or her contribution to the situation, the learners and their contribution need to be described (Allright, 1988). Furthermore, in order to observe the classroom and what goes on in it for the purpose of continued learning and exploration, it is essential to capture the events of the classroom as accurately and objectively as possible and not only to make a record of impressions (Allright, 1988; Wajnryb, 1992). In this respect observation can be more global or more focused. The importance of observation procedures in addition to coding schemes has been emphasized in recent research (Cohen, Manion & Morrison, 2000; Mackey & Gass, 2005).

Sadly, the reality in schools is that over the years observation and the idea of observation has become synonymous with evaluation and punitive action. The potential to create cultures of continuous improvement through observation has stagnated as teachers have felt observations were only for punitive or evaluative reasons. With the concept of the observation stigmatized, other creative ways of observing lessons such as peer observations (by teachers and for teachers) have been limited.

Thus, the researcher came up with this study to develop and validate a classroom observation tool that will be utilized by the different program chairpersons in the university during the delivery of classroom instruction.

Objective of the Study

The purpose of this study is to develop and validate a classroom observation tool to be used by the chairperson and other key officials in the university in observing classes.

Related Literature

Many observation protocols used in both K–12 and postsecondary settings are designed to evaluate the quality of classroom teaching. This judgment can take the form of underlying scales that purport to capture key aspects of high-quality instruction, which are based on either external criteria or latent variables viewed as proxy measures for student achievement. This approach is attractive for those wanting a single measure of teaching quality, or to determine whether or not teaching measures up to particular standards or expectations (e.g., degree of “reformed-ness”). Thus, in a single measure, analysts can determine whether or not a teachers’ practice is aligned with the best practices, as specified by particular criteria (Sawada et al, 2002). Yet caution should be exercised in the use of evaluative protocols for two reasons. First, evaluative measures may turn off or alienate faculty and thus may be of limited utility for professional development purposes. Second, reliability is difficult to attain when analysts are required to not only describe teaching but also judge its quality in real time. A recent review of reliability of evaluative protocols used in K–12 settings found that ratings varied considerably (Guarino & Stacy, 2012), and that rater bias (i.e., pre-existing beliefs about what constitutes high-quality teaching) is

a major reason for the high degree of variability observed in the use of these protocols (Cash, Hamre, Pianta, & Meyers, 2012).

Classroom observation instruments usually focus on measuring specific interactions between students and teachers in the classroom (Board, 2011; Waxman & Huang, 1999). The teaching quality in the classroom that is measured and calculated could be both the teacher's performance and the classroom's effects in response to the complexity of classrooms (Berliner, 2014). However, when classroom observation ratings are used as standard-based evaluations regarding teachers, teachers may need to be detangled from classrooms as teaching contexts. If a teacher's observation ratings are constant across all the classrooms he or she teaches, ratings from any of the classrooms can be representative of his or her teaching performance for personnel decisions (Bell et al., 2012; Kane, 2006). However, if a teacher's observation ratings are not constant across his or her classrooms, measures of teacher effectiveness from a single classroom may not be appropriate for making operational decisions regarding teachers (Kennedy, 2010)

School climate has been associated with many important and beneficial school outcomes. Angell (1991) suggested that positive classroom climate is associated with developing civic virtues in students. Schools with better student perceptions of the teaching climate were correlated with lower student dropout rates by students in their senior year (Barile et al., 2012). Brackett, Reyes, Rivers, Elbertson, and Salovey, (2011) found that student conduct and teacher affiliation were better in schools that scored higher in perceived classroom emotional climate. School climate, specifically school connective health and academic outcomes, violence prevention, student risk behaviors, including sex, violence, and drug use (Cohen, McCabe, Michelli, & Pickeral, 2009). Zullig, Huebner, and Patton (2011) found that school climate was positively associated with student satisfaction and school safety.

Most significantly, school climate is positively correlated with student learning and student motivation to learn, which is ultimately a primary mission for any school (Cohen, McCabe, Michelli, & Pickeral, 2009). Decades of research have clearly established that classroom climate is a predictive variable for student achievement (Arter, 1989; Fraser, 1991; Thapa et al., 2013). Establishing and maintaining a positive classroom climate is basic to improving schools (Adelman & Taylor, 2005).

Technology has provided multiple ways to communicate both in and outside of the classroom, influencing the relationship domain. As Hakkarainen, Muukonen, and Lipponen (2001) state, "Relationships in modern societies are transformed by emerging new means of creating, processing, accessing, and transferring information". This transformation is changing student access to teachers, how students converse with faculty and their peers, and how faculty members collaborate.

Methodology

A. Development Phase

Design

For the development of the new instrument, the approach described by Lynn (1986) was followed. This approach advocates stages. Stage I (development) results in the generation of the instrument's items which was lifted from the literature reviewed and from the existing instruments from the university and other higher teacher education institutions.

B. Validation Phase

Design

For the content validity of the new instrument, the approach described by Lynn (1986) was followed. This approach advocates stage II which intend to evaluates the performance of the instrument's items (judgement and quantification) performed by the content experts which includes the school dean, university psychometrician, program chairperson, and Faculty Federation President.

Locale

The setting of the study was the School of Teacher Education of Biliran Province State University, Naval, Biliran.

Research Respondents

The respondents of the study involved the school dean, university psychometrician, program chairperson, and Faculty Federation President.

Research Instrument

The research instrument were lifted from the existing classroom observation tool of the university and various authors from other learning institutions concerning on classroom observations. The instrument consists of three indicators namely: classroom preparation and management; teaching process and learning engagement; and learners' assessment and reporting.

Data Gathering Procedure

Prior to the final crafting of the instrument to be fielded, the researcher made constant consultation with the research professor and other experts from the university. After such, letter of permission were prepared to formally administer the instrument to the panel of experts. Finally, data were tabulated and interpreted to answer the objective of the study.

Data Analysis

After the retrieval of the instruments from the experts, computation of content validity index of each item were performed. After determining the CVIs, it was interpreted based on the standard as stipulated in table 2. Thus, the researcher made implications based on the results.

Table 1. The scoring method of experts

Relevance	Clarity
1 [not relevant]	1 [not clear]
2 [item need some revision]	2 [item need some revision]
3 [relevant but need minor revision]	3 [clear but need minor revision]
4 [very relevant]	4 [very clear]

To obtain content validity index for relevance and clarity of each item (I-CVIs, the number of those judging the item as relevant or clear (rating 3 or 4) was divided by the number of content experts.

Table 2. The number of experts and its implication on the acceptable cut-off score

Number of Experts	Acceptable CVI Values	Source of Recommendation
Two experts	At least 0.80	Davis (1992)
Three to five experts	Should be 1	Polit & Beck (2006), Polit et al., (2007)
At least six experts	At least 0.83	Polit & Beck (2006), Polit et al., (2007)
Six to eight experts	At least 0.83	Lynn (1986)
At least nine experts	At least 0.78	Lynn (1986)

Table 2 summarizes the recommended number of experts with its implication on the acceptable cut-off score of CVI. It can be concurred that for content validation, the minimum acceptable expert number is two, however most of recommendations propose a minimum of six experts. Considering the recommendations, the number of experts for content validation should be at least 6 and does not exceed 10.

Ethical Considerations

Ethical considerations like anonymity, autonomy, confidentiality, and informed consent were strictly observed during the conduct of the study. Having said this, the respondents were made aware of the purpose of the study. They were treated as more important than the research itself; and therefore, have always been respected. The convenience of the participants was also considered, especially in the current setting. Data gathered were treated as confidential and purely for research purposes only.

Results and Discussion

Table 3 presents the results of the content validity index conducted to determine the validity of the instrument in assessing the performance of the teachers during the teaching and learning process.

Table 3. Content Validity Index on Relevance of the Items

Item No.	I-CVI	Interpretation
1	1	Appropriate
2	1	Appropriate
3	1	Appropriate
4	1	Appropriate
5	1	Appropriate
6	1	Appropriate
7	0.75	Eliminated
8	1	Appropriate
9	1	Appropriate
10	0.75	Need revision
11	1	Appropriate
12	1	Appropriate
13	1	Appropriate
14	1	Appropriate
15	1	Appropriate
16	0.75	Need revision
17	1	Appropriate
18	1	Appropriate
19	0.50	Eliminated
20	1	Appropriate
21	1	Appropriate
22	1	Appropriate
23	0.75	Eliminated
24	0.75	Eliminated
25	1	Appropriate
26	1	Appropriate
27	1	Appropriate
28	1	Appropriate
29	0.75	Eliminated
30	1	Appropriate
31	1	Appropriate
32	1	Appropriate
33	1	Appropriate
34	1	Appropriate
35	1	Appropriate
36	0.75	Eliminated
37	1	Appropriate
38	1	Appropriate

39	1	Appropriate
40	1	Appropriate
S-CVI=0.944		

As shown in table 3, 32 items were rated 1 interpreted as appropriate and 8 items were rated 0.50 to 0.75 which need revision. This indicates that majority of the items developed are relevant items that would serve as tool in observing classes.

Table 4. Content Validity Index on Clarity of the Items

Item No.	I-CVI	Interpretation
1	1	Appropriate
2	1	Appropriate
3	1	Appropriate
4	1	Appropriate
5	1	Appropriate
6	0.75	Eliminated
7	1	Appropriate
8	1	Appropriate
9	1	Appropriate
10	0.75	Eliminate
11	0.75	Eliminate
12	1	Appropriate
13	1	Appropriate
14	1	Appropriate
15	1	Appropriate
16	1	Appropriate
17	1	Appropriate
18	1	Appropriate
19	0.50	Eliminated
20	1	Appropriate
21	1	Appropriate
22	1	Appropriate
23	0.75	Eliminated
24	0.75	Eliminated
25	1	Appropriate
26	1	Appropriate
27	1	Appropriate
28	1	Appropriate

29	0.75	Need revision
30	1	Appropriate
31	1	Appropriate
32	1	Appropriate
33	1	Appropriate
34	1	Appropriate
35	1	Appropriate
36	0.75	Eliminated
37	0.75	Eliminated
38	1	Appropriate
39	1	Appropriate
40	1	Appropriate

S-CVI=0.938

As shown in table 4, 31 items were rated 1 inferred as appropriate and 9 items were rated 0.50 to 0.75 which need revision. Results imply that items developed intended for classroom observation reflects important features of teaching and learning.

Table 5
Reliability of the Test Items

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
A1	145.0833	53.902	.228	.	.935
A2	145.0833	56.265	-.182	.	.939
A3	145.1667	53.606	.233	.	.935
A4	145.1667	52.879	.346	.	.934
A5	145.0000	52.182	.741	.	.930
A6	145.0000	53.273	.475	.	.932
A7	145.0833	51.356	.443	.	.933
A8	145.1667	49.606	.870	.	.927
A9	145.1667	48.152	.790	.	.928
A10	145.0833	50.083	.930	.	.927
B1	145.0833	50.083	.930	.	.927
B2	145.0833	51.720	.622	.	.930
B3	145.0000	54.909	.085	.	.935
B4	145.0000	52.182	.741	.	.930
B6	145.0000	52.182	.741	.	.930
B7	145.4167	50.811	.289	.	.942
B8	145.0000	52.182	.741	.	.930
B13	145.0000	52.182	.741	.	.930
B14	145.4167	53.902	.154	.	.937
B15	145.0833	51.356	.443	.	.933

B17	145.0833	50.083	.930	.	.927
B18	145.0000	52.182	.741	.	.930
B19	145.0833	51.720	.622	.	.930
B20	145.0833	50.083	.930	.	.927
C1	145.0000	52.182	.741	.	.930
C2	145.0000	52.182	.741	.	.930
C3	145.0000	52.182	.741	.	.930
C5	145.0000	52.182	.741	.	.930
C6	145.0833	52.811	.423	.	.933
C8	145.0000	52.182	.741	.	.930
C10	145.0833	50.083	.930	.	.927
Cronbach's Alpha .933	Cronbach's Alpha Based on Standardized Items .952				No. of Items 31

Table 5 shows that the cronbach's alpha is .933 which means that the scale has an excellent internal consistency. It is further noted that 9 items were removed from the scale in the analysis and were considered missing values since each of the items has zero variance. This would mean that out of 40 item scale, there were 31 items remained as part of the observation tool developed.

Conclusion and Recommendation

Content validity is very important to ensure the overall validity of the proposed tool in evaluating performances. Thus, this research endeavor has provided a systematic and evidence-based approach to conduct a proper content validation ensuring that the tool to be utilized in observing classes will give justice and compensate the actual performance of the teachers during the discharge of their duties and responsibilities.

References

1. Adelman, H. S., & Taylor, L. (2005). Classroom climate. In S. W. Lee, P. A. Lowe, & E. Robinson (Eds.), *Encyclopedia of school psychology* (pp. 89-91). Thousand Oaks, CA: Sage
2. Arter, J. A. (1987, April). Assessing school and classroom climate: A consumer's guide. *Resources in Education*, 1-80. Retrieved from ERIC database. (ED295301)
3. Barile, J., Donohue, D., Anthony, E., Baker, A., Weaver, S., & Henrich, C. (2012). Teacher-student relationship climate and school outcomes: Implications for educational policy initiatives. *Journal Of Youth and Adolescence*, 41, 256-267. <http://dx.doi.org/10.1007/s10964-011-9652-8>
4. Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, 17(2-3), 62-87. doi: 10.1080/10627197.2012.715014



5. Berliner, D. C. (2014). Exogenous variables and value-added assessment: A fatal flaw. *Teachers College Record*, 116(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17293>
6. Board, J. (2011). Classroom Observation-Purposes of Classroom Observation, Limitations of Classroom Observation, New Directions. Retrieved from <http://education.stateuniversity.com/pages/1835/Classroom-Observation.html>
7. Brackett, M., Reyes, M., Rivers, S., Elbertson, N., & Salovey, P. (2011). Classroom emotional climate, teacher affiliation, and student conduct. *The Journal of Classroom Interaction*, 46, 27-36. Retrieved from <http://ei.yale.edu/publication/classroom-emotional-climate-teacher-affiliation-and-student-conduct/>
8. Cash, A. H., Hamre, B. K., Pianta, R. C., & Meyers, S. S. (2012). Rater calibration when observational assessment occurs at large scales: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly*, 27(3), 529–542.
9. Cohen, J., McCabe, L., Michelli, N., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111, 180- 213. Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=15220>
10. Fraser, B. J. (1991). Two decades of classroom environment research. In B. J. Fraser & H. J. Walberg (Eds.), *Educational environments: Evaluation, antecedents and consequences* (pp. 3-27). Elmsford, NY: Pergamon Press.
11. Guarino, C., & Stacy, B. (2012). Review of gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. National Educational Policy Center. Boulder, CO.
12. Hakkarainen, K., Muukonen, H., & Lipponen, L. (2001). Teachers' information and communication technology (ICT) skills and practices of using ICT. *Journal Of Technology & Teacher Education*, 9, 181-197
13. Halpin, A. W., & Croft, D. B. (1963). *The organizational climate of schools*. Chicago, IL: Midwest Administration Center, University of Chicago
14. Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger Publishers.
15. Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, 39(8), 591–598. doi: 10.3102/0013189X10390804
16. Lippitt, R., & Gold, M. (2010). Classroom social structure as a mental health problem. *Journal of Social Issues*, 15(1), 40-49. <http://dx.doi.org/10.1111/j.1540-4560.1959.tb01434.x>
17. Perry, A. C. (1908). *The management of a city school*. New York, NY: The Macmillan Company
18. Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83, 357-385. <http://dx.doi.org/10.3102/0034654313483907>
19. Waxman, H. C., & Huang, S. L. (1999). Classroom observation research and the improvement of teaching. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (pp.107-129). Berkeley, CA: McCutchan.
20. Zullig, K. J., Huebner, E. S., & Patton, J. M. (2011). Relationships among school climate domains and school satisfaction. *Psychology in the Schools*, 48, 133- 145. <http://dx.doi.org/10.1002/pits.20532>