International Journal of Education and Development using Information and Communication Technology (IJEDICT), 2016, Vol. 12, Issue 3, pp. 93-107.

Evaluating classroom interaction with the iPad®: An updated Stalling's tool

Gregory MacKinnon and Lourens Schep
Acadia University, Canada

Lisa Lunney Borden, Anne Murray-Orr and Jeff Orr
St. Francis Xavier University, Canada

Paula MacKinnon
MindBloom Consulting, Canada

ABSTRACT

A large study of classrooms in the Caribbean context necessitated the use of a validated classroom observation tool. In practice, the paper-version Stalling's instrument (Stallings & Kaskowitz 1974) presented specific challenges with respect to a) facile data collection and b) qualitative observations of classrooms. In response to these shortfalls, the Stalling's classroom observation tool has since been modified and supplemented with qualitative observation scales. In an effort to promote efficiency of observation and data collection, the tool has been reconfigured as an iPad® application. The research described herein used four independent studies (utilizing video of classroom practice) in an attempt to establish reliability and validity of the new iPad®-based classroom observation instrument. The articulated results provide a good measure that this goal has been reached.

Keywords: classroom observation, technological tools, research instruments

INTRODUCTION: ASSESSING CLASSROOM INTERACTIONS

Classroom interactions have been studied at length from the perspective of both teacher-student interaction and student-student interactions (Cazden & Beck 2003, Fairclough 2013.) Most recently the complex role of technology in mediating learning between teacher and student has also been articulated (Mishra & Koehler 2006, Rosenberg & Koehler, 2015) in the TPACK model. As teachers are encouraged to engage action research in their classrooms as reflective practitioners (Robinson & Lai 2005) quality mixed methodologies for classroom observations have become increasingly important.

The notion of student–centred learning has been promoted for some time through the works of Dewey (1938), Piaget (1977) and Vygotsky (1989). Nonetheless, the ideal of constructivist classrooms (Brooks & Brooks 1999) continues to be hampered by the pressures of standardized assessment (Popham 2001; Ravitch 2011). Widespread assessment trends have been shown (PISA, 2014) to support passive versus active learning. Recognizing the danger of departure from authentic, situated cognition in schools, some Caribbean and Latin American countries have undertaken studies that assess the level of active learning (Vegas, & Petrow 2008). The following research study sought to measure active learning in Barbadian public school classrooms using a valid instrument.

In undertaking the research described herein, a variety of observation tools were considered and discounted for reasons of 1) lengthy and cumbersome recording formats, 2) specialized software required, 3) specialized populations observed and 4) extensive observer training or certification.

The tools considered included: The *Framework for Teaching Evaluation* Instrument created by Charlotte Danielson (2011) utilized by the Bill and Melinda Gates Foundation as one of the instruments in their "Measures of Effective Teaching (MET)" project, Pianta, La Paro & Hamre's (2008) *Classroom Assessment Scoring System* (CLASS) system which requires proprietary software, a lengthy observation guide that accompanies VanTasselBaska, Avery, Struck, Feng, Bracken, Drummond, & Stambaugh's, (2003) *William and Mary Classroom Observation Scales* and a range of population-specific instruments (Cassady, Speirs Neumeister, Adams, Cross, Dixon, & Pierce,2004; Sawada, Turley, Falconer, Benford, & Bloom, 2002; Weiss, Pasley, Smith, Banilower, & Heck, 2003). A simple tool with a manageable learning curve was chosen as best suited for the observation of Barbadian classrooms, the description of which follows.

As early as the mid 1970s, an instrument was designed (Stallings & Kaskowitz 1974; Stallings & Giesin 1977; Stallings 1980) to give a valid measure of active instruction in the classroom. The Stallings Instrument represents a sophisticate three dimensional matrix involving (1) teacher approach, (2) teaching materials used and (3) the size of the teaching and learning groups (i.e. T=teacher, I=student; number of persons 1=single, S=small group, L=large group & E=everyone). This coding instrument was intended to be used in multiple snapshots during a classroom period so as to further differentiate the interaction activity as a function of the class time continuum. For each of 10 snapshots one paper sheet was used to code the teacher and the student activity. The instrument was modified by the World Bank in 2007 to assist in their studies of classrooms in South and Latin America (see: www.eddataglobal.org/embedded/stallings_snapshot.doc). More recently, Bando and Li (2014) have accessed the Stallings tool for a study of teacher training in the context of teaching English as a second language. The grid for scoring classroom interactions is shown in Figure 1. Developers supplemented this instrument with a systematic description of the definitions that scorers would use for assigning appropriate codes. This inherently improved the inter-rater reliability of the instrument.

Inherent Drawbacks of the Stallings Instrument?

A fundamental problem with collecting as many as 10 snapshots per class period in a research study is the amount of paper record it generates and the fact that the data sets must still be collated and added to a spreadsheet in order do any statistical analysis.

In a recent study of classroom practice in Barbados (ATP 2014), researchers found that, not only was the Stallings data collection and associated data manipulation onerous, but that several (1) components of the observation scheme required updating with respect to new classroom tools, (2) the range of observations lacked qualitative overview and (3) the range of technology use by the instructor in the classroom was inadequately measured. Given the nature of these shortfalls, it seemed the application instrument could be significantly improved by considering a technology solution.

Snapshot 1

Classroom Observation Snapshot

ACTIVITY		MATERIALS						
		No Material/None	Books	Notebook/Paper/Pencil	Chalkboard/Blackboard	Manipulative	Visual Aides/Computer/Calculator	Co-operative
1. Reading Aloud	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
2. Instruction/Demonstration/Lect	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
3. Discussion	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
4. Practice/Drill	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
5. Kinesthetic/Projects	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
6. Silent Seatwork	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
7. Copying	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
8. Verbal Instruction	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
9. Other	T	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
	I	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E	1 S L E
10. Social Interaction	T	1 S L E						
	I	1 S L E						
11. Student Uninvolved	I	1 S L E						
12. Discipline	T	1 S L E						
13. Classroom Management	T	1 S L E						
	I	1 S L E						

14. Teacher Social Interaction/Uninvolved [T] **15. Teacher Mgmt.** [T] **16. Teacher out of Room** [T]

Figure 1: *The Stallings paper based instrument. T=teacher, I=student, number of persons 1=single, S=small group, L=large group & E=everyone.*

An Improved Model?

The Stallings instrument was modified by the authors to include not only a broader range of teacher activities, but also an enhanced list of tools the teacher might use in the instruction (see Figure 2a-b). Group sizes were maintained in the instrument (Figure 2c).

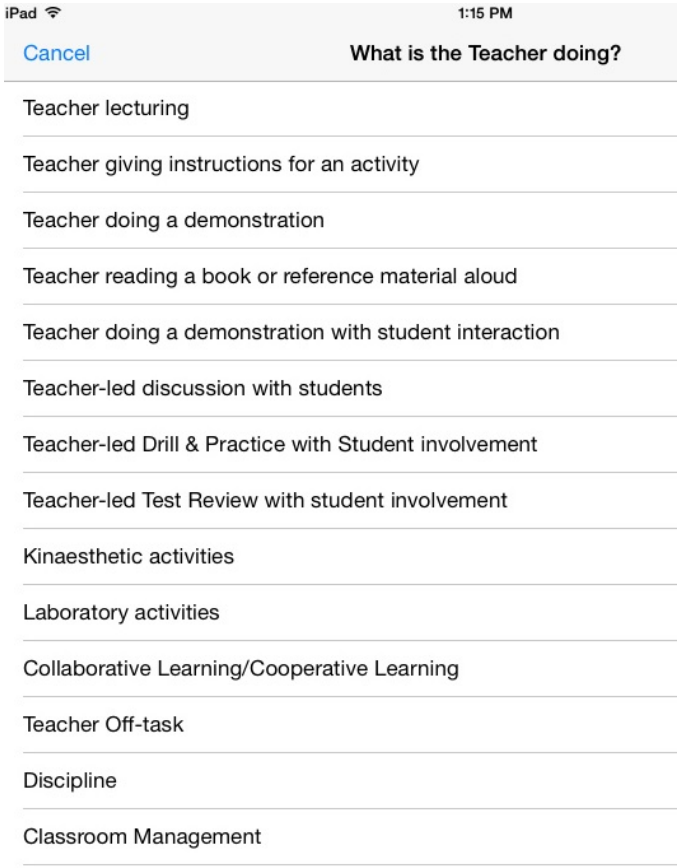

Figure 2a: *iPad® application prompts regarding the teacher's activity*

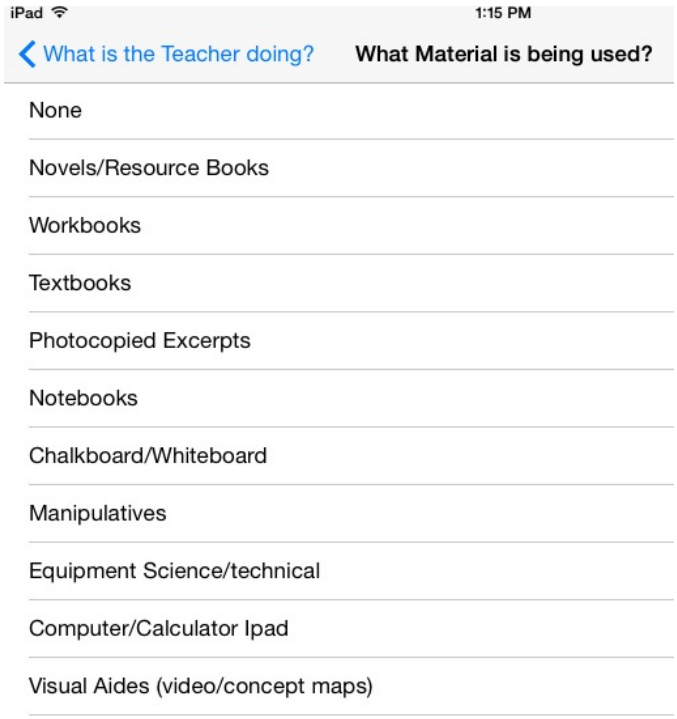

Figure 2b: *iPad® application prompts regarding the teacher's choice of materials*

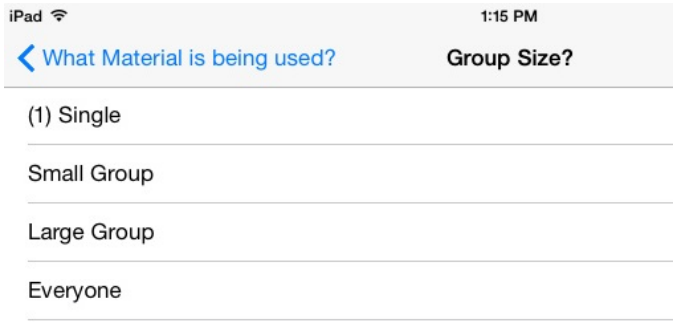

Figure 2c*: The group size that the teacher was working with.*

In response to the lack of qualitative overview of the classroom, a scoring rubric measuring the extent of the following variables was created using observational concepts related to constructs of high quality instruction based on well planned, significant learning through purposeful problem solving and engaged classroom discourse (adapted from McCann, Jones & Aronoff 2012) (see Figure 3).

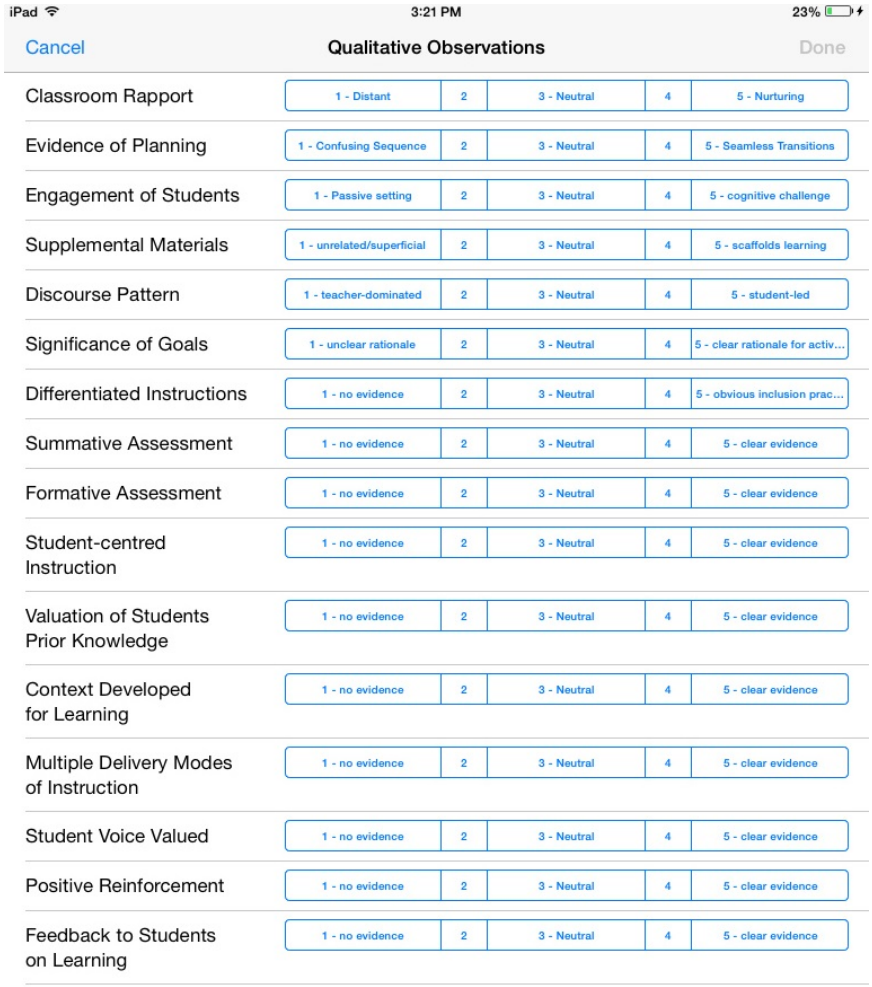

***Figure 3**: Qualitative overview of classroom activity*

THE TECHNOLOGY SOLUTION

The above modifications to the Stallings Instrument were incorporated in the design of an iPad® application. After several design iterations (Willis & Wright 2000), it became evident that a matrix, while visually appealing, was difficult to present legibly in its entirety, on a single screen especially given the extension of variables we had employed. Instead, a field entry model was incorporated where choices about interaction style, tools and group size were selected in sequence during each observation (see Figure 2a-c). After each successive observation, the coder had the option of compiling accumulated data or exiting the loop and recording the supplemental rubric designations for each of the overall qualitative factors (see Figure 3). The application was configured such that a compiled spreadsheet was buffered and automatically uploaded to a database repository (e.g. Dropbox®) the next time the iPad® made an internet connection. It should be noted that all data was tagged because the opening page of the application provided opportunity to enter demographic information including: date, school, grade level, teacher name, observer name, subject, student numbers, gender distribution and length of observation.

VALIDATING THE INSTRUMENT

The iPad® application was studied in several ways in an effort to validate the instrument whilst considering reliability.

Central to investigating the application of this new tool, was the professional preparation of classroom videos. Using two independent videographers in each classroom capture, three separate classrooms were videotaped for their duration of 50 minutes. In the video collection, camera operators were pre-instructed to pan the entire classroom activity including all interactions of teacher and student combinations. The footage from each camera was time stamped and blended to create a single account of the class period. The classroom subjects were purposely chosen to span secondary school mathematics and social studies as well as elementary language arts and science.

Preparing to Evaluate

Four coding evaluators were provided with the rationale for each type of coded interaction and were required to code five short (5 minute) video clips in a practice session (using the categorical questions in Figure 2a-c). A culminating focus group (Morgan, 1997) session with the evaluators helped to clarify any ambiguity in coding assignment. In further preparation, the qualitative Likert component was tested with two 10-minute video clips of classroom interactions; each of the evaluators watched the two clips and assigned Likert values as per Figure 3. While coders were within 80% overlap in their Likert assignment, the individual evaluator's codes were compared for each item and discussed with the group in an effort to promote continuity of assignment (i.e. inter-rater reliability).

Formal Studies

In one study, the three 50 minute videos were watched and coded (using the iPad® app.) by four independent evaluators. The compiled data (in spreadsheet form) was compared across raters (Table 1).

In a second study, 1 month later, the same 4 raters watched the very same 3 videos and coded them again with the iPad® application. This second set of spreadsheet data (Table 2) was compared across raters but also with the first data set to see whether over time there was any change in how each coder may evaluate classroom interaction.

In the third study, each of the four raters was asked to watch the three 50 minute classroom accounts and make field notes about their specific observations and repeat the process using the iPad® application. In subsequent 30-minute audio-recorded interviews, each evaluator was subjected to a standardized open-ended schedule of questions (Patton 1992) that probed (1) the inherent predisposition of the evaluators to using technology, (2) the ease of using the iPad® application and (3) the comprehensiveness of the instrument in capturing the classroom interaction i.e. compared to field notes. Interviews were transcribed and coded for emergent themes (Huberman & Miles 2002) in an iterative axial coding approach (Strauss & Corbin 1991).

In a fourth study, each of the 4 raters were asked to complete a discourse analysis with the iPad® application in 4 separate classrooms at different times. The intent of this approach was to allow the raters to compare qualitatively the difference between coding a video (where the locus of control was in the camera technician) versus coding an actual classroom where they were free to pan the entire group continuously.

The overall findings of the studies were analysed in a constant comparative approach (Glaser & Strauss 1967). Further the results were shared with a colleague (unattached to the studies) in a peer debriefing session (Guba & Lincoln 1994) in an effort to both corroborate analysis approach and discount weak causation relations.

The cumulative results were discussed with the evaluator group in a focus group member-check session (Krueger & Casey 2009; Morgan 1997), to sample consensus on findings and remove spurious outlier results.

RESULTS

Study One

A common qualitative surface comparison for nominal or categorical data is to generate a fraction of the equivalent codes divided by the total number of codes; sometimes expressed as a percentage overlap (Potter & Levine-Donnerstein 1999; MacKinnon, 2003). Consider first the coding (Figures 1-2). In the first study, the average calculation of this indicator was .92 or 92% overlap of coding across raters of the classroom interaction. This rudimentary measure, while suggesting a trend, neglects the "chance" factor of two coders possibly assigning the same codes. We addressed this using SPSS and the Kappa statistic (Hallgren 2012). The Kappa statistic was generated for each paired observation resulting in an overall adjusted average Kappa value of 0.89, a promising indicator of inter-rater reliability of the instrument.

For the qualitative Likert scale (Figure 3), the coding range of each rater for each item is shown in Table 1.

Table 1:. *Qualitative observations of 3 videos by 4 raters (based on Likert scale Figure 3)*

Likert Item	Video 1				Video 2				Video 3			
	Rater 1	Rater 2	Rater 3	Rater 4	R1	R2	R3	R4	R1	R2	R3	R4
Classroom Rapport	4	4	5	4	4	4	4	4	3	4	4	4
Evidence of Planning	4	5	5	5	4	4	4	4	4	4	3	4
Engagement of Students	4	4	5	4	4	4	5	4	3	4	4	4
Supplemental Materials	5	5	5	5	5	5	5	5	5	5	5	5
Discourse Pattern	3	3	3	3	5	5	5	4	4	4	4	4
Significance of Goals	4	4	4	4	3	3	3	4	4	4	4	5
Differentiated Instruction	2	2	2	2	4	4	4	4	3	4	3	4
Summative Assessment	5	5	5	5	5	5	5	5	5	5	5	5
Formative Assessment	4	4	4	4	3	3	3	3	3	4	3	4
Student-centred Instruction	2	2	2	2	4	4	4	5	4	4	4	4
Valuation of Students' Prior Knowledge	4	4	5	4	4	4	4	4	4	4	4	4
Context Developed for Learning	4	5	4	4	4	4	4	3	3	4	4	3
Multiple Delivery Modes of Instruction	5	5	5	5	5	5	5	4	4	5	5	4
Student Voice Valued	4	4	4	4	3	3	3	3	3	3	4	3
Positive Reinforcement	4	4	4	4	4	4	4	4	5	4	5	4
Feedback to Students on Learning	4	3	4	3	4	4	5	4	4	3	4	4

By observation, it is evident that there is not a single incidence where raters vary in their Likert scale assignment by more than one unit. Further, in many cases they are coding their qualitative observations of the videos identically.

Study Two

After one month the 4 raters were asked to watch the same three classroom videos and assign both interaction codes (Figure 1-2) and qualitative observations codes (Figure 3). As per the aforementioned approach, it was found that the adjusted Kappa statistic yielded a .86 or 86% overlap in the coding. Table 2 documents the post data where the numbers in brackets indicate the change from the first coding. With regard to the qualitative coding (Figure 3), again the coding of individual items in no case varies by more than one Likert scale item between raters. Table 3 lists the change in the overall average code for each item. In these averages it is noted that there is no more than 0.25 difference from the first coding to the second coding. It would seem that this indicates that a passage of time has negligible impact on the codes assigned for the same observed classrooms.

***Table 2**: Qualitative observations of 3 videos by 4 raters after 1 month*

Likert Item	Video 1				Video 2				Video 3			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
Classroom Rapport	4(+1)	4	4	4	4	4	4	4	3(-1)	4	4	4
Evidence of Planning	4	5(+1)	5	5	4	4	4	4	4	4	3	4
Engagement of Students	4	4	5(+1)	4	4	4	5	4	3	4	4	4
Supplemental Materials	5	5	5	5	5	5	5	5	5	5	5	5
Discourse Pattern	3	3	3	3	5	5	5	4(-1)	4	4	4	4
Significance of Goals	4	4	4	4	3	3	3	4(+1)	4	4	4	5
Differentiated Instruction	2	2	2	2	4	4	4	4	3	4	3	4
Summative Assessment	5	5	5	5	5	5	5	5	5	5	5	5
Formative Assessment	4	4	4	4	3	3	3	3	3	4	3	4
Student-centred Instruction	2(-1)	2	2	2	4	4	4	5(+1)	4	4	4	4
Valuation of Students' Prior Knowledge	4	4	5	4	4	4	4	4	4	4	4	4
Context Developed for Learning	4	5	4	4	4	4	4	3(-1)	3	4	4	3
Multiple Delivery Modes of Instruction	5	5	5	5	5	5	5	4	4(-1)	5	5	4
Student Voice Valued	4	4	4	4	3	3	3	3	3	3	4(+1)	3
Positive Reinforcement	4	4	4	4	4	4	4	4	5(+1)	4	5	4
Feedback to Students on Learning	4	3	4	3(-1)	4	4	5	4	4	3(-1)	4	4

***Table 3**: Average qualitative codes assigned by raters between first and second coding*

Likert Item	First Coding			Second Coding		
	Video 1	Video 2	Video 3	Video 1	Video 2	Video 3
Classroom Rapport	3.75	4	4	4	4	3.75
Evidence of Planning	4.5	4	3.75	4.75	4	3.75
Engagement of Students	4	4.25	3.75	4.25	4.25	3.75
Supplemental Materials	5	5	5	5	5	5
Discourse Pattern	3	5	4	3	4.75	4
Significance of Goals	4	3	4.25	4	3.25	4.25
Differentiated Instruction	2	4	3.5	2	4	3.5
Summative Assessment	5	5	5	5	5	5
Formative Assessment	4	3	3.5	4	3	3.5
Student-centred Instruction	2.25	4	4	2	4.25	4
Valuation of Students' Prior Knowledge	4.25	4	4	4.25	4	4
Context Developed for Learning	4.25	4	3.5	4.25	3.75	3.5
Multiple Delivery Modes of Instruction	5	4.75	4.75	5	4.75	4.5
Student Voice Valued	4	3	3	4	3	3.25
Positive Reinforcement	4	4	4.25	4	4	4.5
Feedback to Students on Learning	3.75	4.25	4	3.5	4.25	3.75

Study Three

Four raters were asked to watch three 50 minute classroom videos and make field notes about what they saw. The raters were teachers with 10, 12, 15, and 20 years teaching experience. Respectively, one was a secondary math educator, one was a secondary social studies educator, one was an elementary language arts educator and one was an elementary science educator. They were then asked to watch the same videos again and use only the iPad® application.

Study Four

Each of the 4 raters were asked to complete a discourse analysis with the iPad® application in a separate face to face classroom.

Following the 4 independent studies, each rater was interviewed for 30 minutes using a standardized open-ended interview schedule. The objective of the interviews was to determine: (1) the inherent predisposition of the evaluators to using technology, (2) the perceived ease of using the iPad® application and (3) the comprehensiveness of the instrument in capturing the classroom interaction i.e. compared to field notes.

The first group of questions posed to these raters was concerning their predisposition to technology. The rationale was to establish whether a possible negative view of technology may in fact impede the open-mindedness of participants to the potentials of technology. This approach was meant to set a baseline such that observed attitudes could be attributed directly to the iPad® application and not other technology experiences (good or bad).

All raters were found to possess a healthy critical attitude towards technology. A representative response was “I have used technology in my teaching and appreciate how it can accomplish certain tasks yet I am quick to abandon non-productive approaches; it is just another tool at my disposal.”

The four raters were asked about the ease with which they used the iPad® application. They were unanimous in suggesting that the application was easy to use yet, all qualified their response by saying that it was extremely important to spend sufficient time doing trial runs where they learned about the definitions of the codes. This was not a surprising finding given that the use of the Stallings instrument, even in its paper format, has been predicated on a systematic preamble (see: www.eddataglobal.org/embedded/stallings_snapshot.doc).

Interviewees found the process of taking field notes of their classroom videos instructive. The 3 most common observations were (1) it is difficult to write and watch what is going on as the video camera pans the classroom, (2) a coding system accessing either paper or using technology definitely streamlines the process and (3) that watching a video and coding is different than coding in an actual classroom; a classroom observation is more difficult. An extension of this was another commonly held introspective expressed in the following statement. “It is predictable with any classroom coding system that raters in real classrooms may, through distractions or inherent viewing preferences, observe slightly different or more complete interactions especially compared to our video observations”

The primary rationale for having raters go back and use the iPad® application to make observations of the videos was to engage a discussion of the validity of the instrument. Interviewees were asked how the cumulative iPad® snapshots compared with their field notes. The following are representative comments.

> “I think that overall I get a sense of the type of classroom by sitting and taking notes, but the application captures that better because it forces me to articulate the components of classroom interaction that come together to make either a passive or active setting” (rater 2)
>
> “The application is interesting because it shows you (through the repeated snapshots) how the classroom interaction changes over the duration of the class; I think the iPad® app collects the data that supports an overall conclusion around student-centeredness” (rater 1)
>
> “I liked the qualitative scales because they went beyond the statistical recording of the number and type of teacher –student interactions to make a statement about the whole class experience- this is the part that overlaps with my own field notes. The beauty of course in the initial interaction codes is that it gives you tangible evidence of what you are seeing overall.” (rater 4)
>
> “The iPad® program generates data that is not only complementary to what I have seen (and taken field notes on) but much more complete in terms of giving me specific examples of the types of interactions at certain times in the classroom” (rater 2)
>
> “The codes for interaction between teacher and student have evolved from the initial Stallings work but the addition of the qualitative Likert scale makes me feel much more confident in the instrument’s ability to speak to the entire classroom experience.”(rater 3)
>
> “The coding of interactions by iPad® app at time intervals is much more efficient at systematically sampling individual interactions then I could be in taking notes. Having said that, the overall assessment of whether the classroom is constructivist by nature comes

from watching the whole class. The app gives you everything though, the data and the qualitative overview." (rater 4)

"I like the way the Likert scale forces you to deconstruct the pieces that make the classroom student-centred" (rater 1)

In the focus group, the 4 raters were asked pointedly whether they felt the iPad® version of the Stalling's instrument captured the essence of what was happening in the classroom videos. In paraphrased form, they posited the following.

- The instrument effectively generated both quantitative and qualitative indicators of what was happening in the classroom both with respect to interactions and overall ""nature" of the classroom experience.
- By comparison to field notes, there was much more data to draw on to support an analysis of the classroom discourse.
- The iPad® version is very easy to use and much less onerous than the paper-based approach in terms of collecting and collating data.
- The addition of the qualitative observations (through the Likert scale) by comparison to the earlier Stalling's instrument (Stallings & Kaskowitz, 1974; Stallings & Giesin, 1977) provides an extended comprehensive overview of the classroom discourse.
- The coding of interactions between teacher and student and the subsequent materials used is very useful in determining the overall nature of the classroom, whether it be an inherently passive or active environment. However, the inter-rater reliability is arguably improved by reviewing the code definitions through a significant practice coding period and concomitant discussions.

Further Work

The raters suggested that the definitions for both interaction codes (Figure 1-2) and Likert scales (Figure 3) be revisited and further clarified to remove ambiguity. Given that this was the first edition of the iPad® application, raters saw room for extending the demographic information that could be entered. When asked whether the qualitative Likert scales should have additional items, the focus group felt that "rater fatigue" may result from too many items; in fact the group suggested we consider the overlap of some items and perhaps collapse similar item into more general categories. One rater suggested that with the current research focus on the potential for technology to enhance learning, the iPad® application might assess the nature of technology use through the popular technology matrix (see: http://fcit.usf.edu/matrix/matrix.php). This tool is currently available electronically but would require appropriate permissions to add to our iPad® application.

While the pilot study of this new iPad® application shows promise, it is only in repeated classroom research use that elements will modified and improved in later editions. Access to trials of this discourse tool can be arranged through MindBloom Educational Consulting (MC, 2016).

REFERENCES

ATP–Analysis of Teaching Practices 2014, *Inter-American Development Bank Project*-Barbados: Internal Publication.

Bando, R. & Li, X. 2014, "The effect of in-service teacher training on student learning of English as a second language". Washington: *Inter-American Development Bank*. Retrieved from http://publications.iadb.org/bitstream/handle/11319/6596/int4FE4.PDF;jsessionid=8AE3142EA90163C08FF7A38D29EA8C53?sequence=1

Brooks, J. G., & Brooks, M. G. 1999, "*In search of understanding: The case for constructivist classrooms"*. Alexandria, Virginia: ASCD.

Cassady, J. C., Speirs Neumeister, K. L., Adams, C. M., Cross, T. L., Dixon, F. A., & Pierce, R. L. (2004) The differentiated classroom observation scale. *Roeper Review*, vol. 26, no, 3, pp.139-147.

Cazden, C. B., & Beck, S. W. 2003, "Classroom discourse. *Handbook of discourse processes*" pp.165-197. New York: Routledge

Danielson, C. (2011). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved from: http://ok.gov/sde/sites/ok.gov.sde/files/TLE-DanielsonFramework.pdf

Dewey, J. 1938, "*Experience and education".* New York, NY: Collier Macmillan.

Fairclough, N. 2013, "*Critical discourse analysis: The critical study of language"*. New York: Routledge

Glaser, B. & Strauss, A. 1967, "*The discovery of grounded theory: Strategies for qualitative research"*. New York: Aldine De Gruyter

Guba, E. G. & Lincoln, Y. S. 1994, "Competing paradigms in qualitative research", in: N. K. Densin & Y. S. Lincoln (Eds) *Handbook of qualitative research* (London, Sage).

Hallgren, K. A. 2012, "Computing inter-rater reliability for observational data: An overview and tutorial". *Tutorials in quantitative methods for psychology*, vol. 8, no.1, pp. 23-34.

Huberman, A. M., & Miles, M. B. 2002, "*The qualitative researcher's companion"*. London: Sage.

Krueger, R. A., & Casey, M. A. 2009, "*Focus groups: A practical guide for applied research"*. Thousand Oaks: Sage.

MacKinnon, G. R. (2003). Inter-rater reliability of an electronic coding system. *Technology, Pedagogy and Education 12*(2), 219-230.

MC (2016). MindBloom Educational Consulting can be reached via paula.mackinnon@mindbloom.ca

McCann, T. M., Jones, A.C., & Aronoff, G.A. 2012, "*Teaching matters most: A school leader's guide to improving classroom instruction".* Thousand Oaks: Corwin.

Mishra, P., & Koehler, M. J. 2006, "Technological pedagogical content knowledge: A framework for teacher knowledge". *Teachers College Record*, vol. 108, no. 6, pp.1017–1054

Morgan, D. 1997, "*Focus groups as qualitative research* (2nd ed.)". Thousand Oaks, CA: Sage.

Piaget, J. 1977, "*The development of thought: Equilibrium of cognitive structures"*. New York, NY: Viking Press.

Pianta, R.C., La Paro, K., & Hamre, B. K. (2008) *Classroom Assessment Scoring System*. Baltimore: Paul H. Brookes. Retrieved from: http://www.teachstone.org/about-the-class/

PISA (Program for International Student Assessment) Retrieved from http://nces.ed.gov/surveys/pisa/

Popham, W. J. 2001, "*The truth about testing: An educator's call to action"*. Alexandria, Virginia: ASCD.

Potter, W. J., & Levine - Donnerstein, D. 1999, "Rethinking validity and reliability in content analysis". *Journal of Applied Communication Research*, vol. 27, no. 3, pp. 258

Ravitch, D. 2011, "*The death and life of the great American school system: How testing and choice are undermining education"*. New York: Basic Books.

Robinson, V., & Lai, M. K. 2005, "*Practitioner research for educators: A guide to improving classrooms and schools"*. Thousand Oaks, California: Corwin Press.

Rosenberg, J. M., & Koehler, M. J. 2015, "Context and technological pedagogical content knowledge (TPACK): a systematic review". *Journal of Research on Technology in Education*, vol. *47*, no. 3, pp. 186-210.

Sawada, D., Turley, J., Falconer, K., Benford, R., & Bloom I., (2002). Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol, *School Science & Mathematics*, vol.102, no. 6, p245

Stallings, J. & Giesen, P. 1977, "The Study of Reliability in Observational Data". Phi Delta Kappa, Occasional Paper 19.

Stallings, J., & Kaskowitz, D. 1974, "*Follow through classroom observation evaluation, 1972-1973"*. Menlo Park, CA: Stanford Research Institute.

Stallings, J. 1980, "Allocated academic learning time revisited, or beyond time on task". *Educational Researcher, vol.* 9, pp. 11–16.

Strauss, A. & Corbin, J. 1990, "*Basics of qualitative research: Grounded theory procedures and techniques"*. Newbury Park, CA: Sage

VanTasselBaska, J., Avery, L., Struck, J., Feng, A., Bracken, B., Drummond, D., & Stambaugh, T. (2003). *The William and Mary classroom observation scale–revised*. Williamsburg, VA: The College of William and Mary
Retrieved from:
http://education.wm.edu/centers/cfge/_documents/research/athena/cosrform.pdf

Vegas, E., & Petrow, J. 2008, "*Raising student learning in Latin America: The challenge for the 21st century"*. Washington: World Bank Publications.

Vygotsky, L. 1989, "*Thought and language"*. Cambridge, MA: MIT Press.

Weiss, I.R., Pasley, J.D., Smith, P.S., Banilower, E.R., & Heck, D.J. (2003). *Study of K-12 mathematics and science education in the United States*. Horizon Research, Inc.

Willis, J., & Wright, K. E. 2000, "A general set of procedures for constructivist instructional design: The new R2D2 model". *Educational Technology*, vol. 40, no. 2, pp. 5-20.