Challenges, Benefits, and Considerations When Conducting Classroom Video Observation Research

Contributors: Leigh McLean & Carol McDonald Connor
Pub. Date: 2018
Access Date: January 16, 2018
Academic Level: Postgraduate
Publishing Company: SAGE Publications Ltd
City: London
Online ISBN: 9781526436252
DOI: http://dx.doi.org/10.4135/9781526436252

Abstract

Classroom observation research using video data has the potential to provide rich, "high-yield" opportunities to investigate important classroom processes that affect student development. However, collecting, assessing, and analyzing classroom observation data come with a unique set of challenges. Furthermore, the successful creation and application of classroom observation tools take careful consideration by investigators. In this case study, we use a published piece and the larger longitudinal study from which this piece stems to guide a discussion of classroom video observation methods. This discussion includes important considerations when developing novel observational tools, as well as investigator recommendations for how to collect, assess, and analyze video observation data. With this case study, we hope to provide practical advice for those hoping to conduct high-quality classroom observation research, with the goal of ensuring that the field continues to move forward through positive contributions involving these methods.

Learning Outcomes

By the end of this case, students should be able to

- Understand the difference between classroom- and student-level observation methods and between global, frequency, and duration observation tools
- Articulate a plan for how to approach the development and refinement of a new classroom observation tool
- Articulate the common challenges, and solutions to these challenges, faced when conducting classroom video observations
- Understand the processes involved in assessing video data using classroom observation tools, including observer training and inter-rater reliability
- Understand the important considerations involved in the accurate statistical analysis of classroom observation research, including identifying the extent to which data are nested and accounting for variance at multiple levels in analyses

---

Project Overview and Context

The study of focus, entitled “Classroom quality as a predictor of first graders’ time in non-instructional activities and literacy achievement” (McLean, Sparapani, Toste & Connor, 2016), utilized both classroom and student-level observation methods to investigate how the quality of first-grade classrooms influenced how students spent their time during instruction, and how these factors contributed to their literacy achievement. We investigated two types of instructional experiences: students’ off-task behavior and time in transitions between activities.

Off-task behavior was defined as any time a student was observed to be engaging in something other than the intended learning opportunity, and transitions were defined as teacher-sanctioned (and usually teacher-initiated) movement between learning opportunities. What we found was that higher classroom quality was associated with less initial student time off-task, and to decreasing time across the year. In addition, higher classroom quality was related to more initial student time in transitions, and to decreasing time across the year. These patterns of change across the year off-task and in transitions mediated a relation between classroom quality and students' literacy achievement.

This investigation was one of many carried out using data from the Individualizing Student Instruction (ISI) longitudinal project (Connor et al., 2013), which began in 2005-2006 and followed more than 3,000 students in multiple cohorts from kindergarten through fifth grade. This project had a heavy classroom observation component, which included collecting whole-day video recordings for each participating classroom at three time-points across each year and later assessing these videos using multiple investigator-developed observational tools. The ISI project represented important progress in the field as it enabled investigators to create, apply, and refine new tools for classroom observation that have allowed us to answer important questions about what students experience in the classroom and how these experiences might influence their development.

The two observational tools used in the study of focus were the Quality of the Classroom Learning Environment rubric (Q-CLE; Connor et al., 2014) and the ISI observational system (Connor et al., 2009). Each of these reflects a different level of classroom observation; the Q-CLE observes the classroom as a whole, or at the classroom level, whereas the ISI system tracks the experiences of each student at the individual, or student level. The Q-CLE assesses overall classroom quality by simultaneously considering the quality of instruction, level of classroom organization, and warmth/responsiveness/control/discipline of the teacher. This measure offers users a way to gain insight into the classroom systems that all students are regularly exposed to. The ISI system, in contrast, tracks the amount of time individual students spend in various instructional and non-instructional activities throughout the day. This system provides a level of detail not often achieved by other student-level observation tools: Each student activity is assessed across multiple dimensions, including the subject being taught, the content being presented, who (the teacher vs. the student) is directing the learning opportunity, and how students are grouped in the classroom in relation to the teacher and their peers.

Student-level observational tools such as the ISI system bring to the table a unique ability to describe much more specifically what is happening within the classroom compared with the "global average" captured by classroom-level tools, a crucial perspective when considering that students within the same classroom often have vastly different experiences (Connor et al., 2009). When used together as in the study of focus, the Q-CLE and the ISI tools provide views of both the larger systems present within the classroom (the three components of classroom quality) and the variation between students in the more specific experiences they have across a day of instruction. This combined view allows us as investigators to answer some very interesting questions about how development takes place within the classroom context, such as those posed in the study of focus.

Classroom observation methods provide investigators with a way to assess students' classroom experiences and draw important inferences about what types of experiences are more or less effective in promoting successful learning. However, past work has shown that there is limited accuracy when outside observers attempt to judge a teacher's/classroom's effectiveness (Strong, Gargani, & Hacifazlioğlu, 2011). This indicates a need for more high-quality, validated tools of observation that make more explicit what to look for when observing a classroom and the students therein. Informed by our experiences with the ISI longitudinal study and resulting investigations, we hope to provide some guidance on how to collect classroom video observation data, how to use these data to answer your research questions, and how to develop, refine, and apply high-quality observational tools that can advance the field's ability to understand classrooms.

Research Design

The ISI project began in 2005-2006 and followed more than 3,000 students in multiple cohorts across elementary school. The portion of this sample used for the study of focus were 533 first-grade students and their 49 teachers who were assessed and observed in the 2007-2008 academic year. Data collection among this sample took place at three time-points across the academic year: once each in the fall (September/October), winter (December/January), and spring (April/May). Student academic achievement data, as well as whole-day video observations, were collected at each time point. The designated literacy blocks of each whole-day video observation were later assessed by trained research assistants using both the Q-CLE and the ISI systems. Mediation analysis was conducted using hierarchical linear modeling (HLM) to account for the nested structure of the data.

Method in Action and Practical Lessons Learned

Creating Classroom Observational Tools

Although much has been accomplished in the area of classroom observation research, many questions remain about the dynamic nature of the classroom and how this environment shapes the outcomes of students. A key element of the study of focus was the application of two novel, investigator-developed classroom observation tools, one at the classroom level (Q-CLE) and one at the student level (ISI). The creation, application, and refinement of new observational tools such as these is a multistep process, and many factors must be attended to before a new observational tool can be used to answer research questions. Following we make some recommendations to aid those who are seeking to create their own classroom observation tools.

First, investigators should approach the creation of a new tool with a clear idea of what they seek to observe, how they will observe it, and a foundational theory with which they have aligned to inform that tool's creation. Knowing these things before designing the tool itself will ensure clarity throughout the entire process and will result in a tool that is not only more reliable in detecting its intended classroom features but is also well received by reviewers when the time comes for this tool to be published. For example, before the official manual for the Q-CLE was written, we had a firm idea of exactly what we wanted this measure to capture: the Q-CLE seeks to capture overall classroom quality in a way that incorporates the interactions and instruction that take place within the classroom, as well as the physical structure and function of the environment. The Q-CLE accomplishes this by considering three separate yet overlapping classroom dimensions (classroom organization, instructional quality, and teacher warmth/responsiveness/control/discipline) which together are thought to represent overall classroom quality. We approached the creation of this tool relying on Dynamic Systems Theory (Thelen & Smith, 1996; Yoshikawa & Hsueh, 2001) which posits that within a classroom, multiple systems exist which interact with each other to affect students' experiences. In the study of focus, as well as in the original paper in which the Q-CLE was introduced (Connor et al., 2014), we had to provide detailed descriptions and justifications of how we conceptualized classroom quality, how we applied the tool to video data, and the theory that drove these processes. Had we not been able to address those points, it is likely this tool would not have been received well in the peer-review process.

Second, each observational tool should be accompanied by a user manual/guide that clearly describes each feature of the classroom you are intending to measure and how observers should go about assigning any judgments associated with each feature. This aligns with the above point of knowing exactly what you want to measure with your tool, but extends this point in that these expectations and corresponding procedures should be written down in an organized, detailed, and easily accessible manual. Of course, this manual will change as you refine your tool, and so keeping track of each version as you make changes is also important so that you can easily reference past versions and track your decision-making processes as the tool evolves.

Third, investigators should approach the creation of a new tool with a thorough understanding of what makes that tool innovative and what research questions they seek to answer. In other words, how will this tool improve, expand on, or change how classrooms are currently being measured? And will it help answer critical research questions? If this tool fills a gap by measuring something that has never been measured before, as was the case with the ISI system, this piece is fairly straightforward. However, if this tool is taking a new approach to measuring something that has already been investigated, as was the case for the Q-CLE, the added benefit of the new tool must be clear. Using the example of the Q-CLE, multiple measures of classroom quality existed before we developed this measure, but we felt that each of these systems was missing an element or elements of the classroom that we thought were important to consider. Specifically, other systems seemed to hold primary focus either on the instruction taking place within the classroom or on the relationships and interactions among teachers and students. Although these classroom elements are no doubt important, we noticed a gap in considering these things along with the physical structure and function of the classroom (e.g., classroom organization, individualization of instruction), and thus, the Q-CLE extended previous measurements of classroom quality by incorporating all of these elements simultaneously. Thinking in this way and knowing the “story” of your measure in terms of how it adds to the field will not only make this measure more compelling in peer review but, most importantly, will also ensure your tool helps move the field of classroom observation in the right direction through positive, innovative contributions. That is, a new measure should make more clear, rather than less clear, the elements of a classroom that are important to student development by adding to and expanding on what already exists.

Finally, investigators should apply their observational tool to “practice” video data to gain insights into which elements need refinement *before* beginning a formal study. These videos should be of classrooms of a similar grade level but should not be the videos you will use in your primary study. If you do not have video data available for this, public data sets that contain such videos can be accessed (for example the Measures of Effective Teaching data; see Kane, McCaffrey, Miller & Staiger, 2013). Even when strongly grounded in theory and painstakingly planned out, the truth remains that the first draft of an observational tool is still just that: a first draft. It takes applying this tool to actual videos to gain insight into what aspects of the tool seem to be successfully representing their intended classroom features and which may need rethinking. Getting this iterative refinement process done before your formal study will ensure that the eventual measurement of your target classrooms is as accurate as possible.

Conducting Classroom Video Observations

Video data provide investigators with the opportunity to conduct multiple investigations into myriad classroom processes that might influence student (and teacher) outcomes. If captured and stored well, these data can be revisited again and again for new investigations; however, some pitfalls may limit investigators' abilities to get the most out of these data. Based on our experiences collecting and assessing video data during the ISI studies, we have some recommendations for how to capture the best observations possible. The first is to capture the classroom from multiple viewpoints with multiple cameras, rather than just a single camera. For each classroom video captured in the ISI studies, one to two trained videographers managed two cameras simultaneously, with one capturing a wide view of the classroom and the other capturing a closer view of the teacher and nearby students. This approach helped us account for the fact that in the typical elementary classroom, there is a lot going on that is easy to miss with a single viewpoint. Often students would be working in multiple small groups across the whole classroom, with some students out of the sight or audio range of one camera. As such, having multiple simultaneous viewpoints of the classroom ensures that as much information as possible is available during later assessment of video data. This approach also proved especially helpful in supporting the applications of both classroom- and student-level observational tools as it ensured that both the teacher and most, if not all, students were captured on video throughout the day.

Second, we recommend that in addition to video data, videographers gather detailed written information about the students, teacher, and activities taking place throughout the observation. In the study of focus, videographers wrote physical descriptions of each child in the classroom and took written notes of all observed classroom activities in real time, with these notes later being matched up to videos for use during video coding. This allows coders to ensure that target children are identified correctly when conducting later student-level observations and provides coders with important "checkpoints" to refer to when trying to capture the extent of what is going on during instruction, which is at times a daunting task.

Third, we recommend that those videotaping classroom observations be familiar with the observational tool(s) that will later be used to asses video data. Each observational tool assesses different, and often very specific, elements of the classroom environment. As such, having a clear, unobstructed view of those elements in classroom videos is essential for assessors to later make accurate judgments. When a videographer knows exactly which classroom features will be the focus of later assessments, they can make efforts to ensure that those elements are captured clearly in the videos. For example, the ISI system documents in extreme detail the type of instruction that a student is receiving during academic instruction (e.g., specifying that a student is partaking in a lesson on phonemic awareness). This can be particularly difficult for an assessor to deduce when students in the class are, say, rotating in groups among multiple different centers, with each center involving a different type of instructional activity. As such, the choices the videographer makes during videotaping can “make or break” a coder’s later ability to accurately code students during group work. Trained ISI coders know that teachers will usually provide students with a detailed overview of each group activity before group work begins and will also review how groups are to rotate between activities. If an ISI coder is videotaping the classroom, they will know to focus on the teacher and any important reference materials (student group flow charts, activity descriptions) with one of the cameras so that coders can later use these as guides. Later, they will also know to capture a close view of all materials being used at each station so that observers can verify that they are accurately assessing exactly what students are doing. In this case, the videographer’s knowledge of the ISI system allowed them to focus on certain situational features that later helped observers determine which students to focus on, how students were moving through the classroom, and what types of instruction students were receiving. Important to note, however, is that if the videographer is also a later assessor of video data they should not code the same videos that they captured, should not know the research questions the study is attempting to answer, and should not know whether classrooms were in treatment or control conditions for experiments as knowledge of these introduces bias.

Finally, we recommend that those conducting video observations make every attempt to be as unobtrusive as possible when collecting these data. It is no surprise that teachers (and some students) are hesitant when it comes to letting strangers with cameras into their classroom. It has the potential to disrupt instruction (especially in younger classes—cameras are indeed fascinating to young children) and can make individuals uncomfortable when thinking about being watched by strangers at a later time. The goal of any video observation is to capture a view of the classroom that accurately represents what typically takes place within that setting; however, the harsh reality is that simply by conducting an observation, the behaviors of those you are observing might change. Encouragingly, in our experiences, we have found that if videographers are respectful, reassuring, and, most importantly, unobtrusive, teachers and students are more at ease and even report forgetting about the cameras once instruction begins. But what do we mean when we say “unobtrusive”? Specifically, we recommend that videographers try to stay out of “high-traffic” areas as much as possible, speak as little as possible and when they do speak in quite voices, and refrain from interacting with students. A particularly important piece here is refraining from interacting with students. Often, students will ask videographers for help with class material and so on, and it is crucial that videographers do not engage in these instances beyond giving a quiet, friendly reminder that they cannot answer questions. Engaging in these situations would change the nature of the instruction you are attempting to observe.

Applying Observational Tools to Video Data

Once the observational tool has been created and refined, and once video data have been collected, the process of assessing (or coding) video data begins. Conducting reliable video observations requires a number of resources and procedures, all of which are at least partially determined by the type of observational tool you are using. Observational tools such as the Q-CLE that are applied to the entire classroom as a whole, commonly referred to as global observation measures, typically involve the assignment of judgments by observers to various classroom-level elements. For example, in the Q-CLE system, observers assign a single score for each of the three classroom quality dimensions, with scores ranging from 1 (poor) to 6 (exemplary). Procedures are similar for other global observation systems. As such, these tools do not rely heavily on resources such as computer software or scoring booklets. Student-level tools, however, are more complex. These systems track in much greater detail what individual students are doing/experiencing in the classroom and as such require comparably more effort and resources.

Student-level tools generally fall into one of two categories: frequency or duration. Frequency tools track how often a student is observed to do/experience a target behavior or activity, for example, participating in a certain type of instruction. In contrast, a duration tool tracks *how long* a student is observed to be doing/experiencing a target behavior or activity. In a frequency tool, each time a target behavior/activity is detected for an individual student, the observer assigns a single code denoting that the event took place. A frequency tool can have multiple codes representing multiple target behaviors/activities, but the key feature here is that observers merely note when something of interest happens, rather than also considering how long that thing occurs (see McLean & Connor, in press for an example of a novel, frequency-coded student-level observational tool). In duration coding, the length of time students spend participating in target activities is also captured. This entails documenting the start and end times of each activity and using these to calculate the number of minutes/seconds each student spent in each target activity throughout the duration of the observation. Student-level tools that account for frequency or duration are more complex than global observation tools and will likely require the use of resources such as specialized software. For example, in the study of focus, Noldus Observer Pro software (www.noldus.com), a program designed specifically to aid investigators in the student-level assessment of video data, was used. This software provided an interface that allowed observers to view video data while simultaneously assigning codes (including start and end times) for target activities. An alternative may be to create an excel spreadsheet into which an observer enters codes for a video they watch in a separate window, but we strongly advise that investigators use more formal systems (such as Noldus Observer Pro or something comparable) whenever possible.

Another important step in applying observational tools is to ensure that all observers are applying the tool in ways that are both in-line with what the developer/primary investigator intended and are consistent with each other. To accomplish this, two steps must be taken: first, primary investigators should implement a rigorous system of training on the use of the tool with the observers; second, the observational team should achieve adequate inter-rater reliability (kappa > .65) prior to formal analysis of study videos and should adhere to a regular schedule of reconfirming inter-rater reliability throughout the duration of the tool's use.

We recommend a scaffolded approach to training that includes initial introduction to the tool by an expert, group discussion of the tool's use, and finally independent practice. The trainer (ideally the primary investigator and creator(s) of the tool) should introduce the tool to the entire group of observers and provide a thorough explanation of what the tool measures and how to use it, along with the final version of the tool's manual for each observer to reference whenever needed. After initial introduction of the tool, the trainer should provide guided examples with practice video data (again, it is important that these data are not videos from your primary study) in a group setting, showing observers in real time, with real video data, how to use the tool. Once this step is complete, we recommend that the entire group complete the assessment of a different video while engaging in discussion and debate and reaching group consensus for each area of assessment. This step can be repeated with multiple group practice videos until all trainees report feeling comfortable using the tool. Next, each observer should complete an independent assessment of another practice video, and once these are complete, the group should come together to discuss their results and compare/contrast their conclusions. Again, this step can be repeated multiple times until the trainer and trainees are all comfortable performing a formal test of inter-rater reliability.

Once training is complete, it is time to test inter-rater reliability. We recommend randomly selecting 10% to 20% of the primary study videos and having each observer assess this set of videos independently using the observational tool. It is crucial that during this time observers do not discuss the videos being assessed with each other or with the trainer. Upon each observer's completion of the reliability videos, two forms of inter-rater reliability should be calculated: percent-agreement and kappa. Percent-agreement reflects the percentage of matching judgments between coders out of all judgments assigned. Although percent-agreement is generally regarded as a reliable method to asses inter-rater agreement, it has been criticized for its lack of ability to account for instances of chance agreement. To remedy this, inter-rater reliability should also be evaluated using Cohen's kappa (Cohen, 1968). In general, kappa values above .75 are considered ideal and above .65 acceptable. One benefit of using software that has been developed specifically for the use of observational tools (such as Noldus Observer Pro) is that such programs usually come equipped with the ability to calculate kappa among multiple observations. If such a program is unavailable, we recommend finding reliable sources that describe how to calculate kappa accurately. If acceptable levels of percent-agreement and kappa are reached, observers can start the exciting process of formally assessing video data. If adequate reliability is not established, the team should repeat training and try again, going through this process until reliability is established. This will likely include updating the coding manual with greater detail and examples. As observers progress in the analysis of video data, a process that can take a long time depending on the complexity of the system, we recommend adhering to a regular schedule of inter-rater reliability assessment following these same procedures with new videos. In the ISI study, observers completed inter-rater reliability assessments every 3 to 6 months.

Analyzing Observation Data

Once you are ready to investigate your research questions using observational data, there are some things you must attend to in order to ensure your analytic methods are appropriate for what you hope to achieve. First, it is important to determine how best to calculate your variables for analysis. Although this may at first seem fairly straightforward, care must be taken to create your variables in a way that best represents the data. This includes checking and accounting for non-normal distributions in the data; making judgments about whether to use total scores, average scores, or adjusted scores; and using advanced analytic techniques such as factor analysis to inform the creation of factors where needed (again, see McLean & Connor, in press for a discussion of factor analysis with a novel student-level tool). Like all classroom observation systems, practical implementation introduces construct-irrelevant variation or measurement error. For instance, past analysis of the ISI and Q-CLE systems has indicated that in addition to variation stemming from teachers and students, variation frequently arises from differences among, for example, raters, lessons, indicators, classroom settings, and their interactions (Nielsen, 2007). Synthesis of results in the literature suggests that teachers are likely to account for approximately 35% of the variance, lessons for 10%, segments for 20%, raters for 10%, and students for 25% (e.g., Carlisle, Kelcey, & Berebitsky, 2013; MET, 2012)

Observational systems, and especially frequency systems, often result in non-normally distributed data. As such, thorough investigation of these variables prior to their use in formal analyses is crucial as some may need to be adjusted to account for this non-normal distribution. In addition, justification should be provided for decisions to either sum or average observational data to create analytic variables. Furthermore, investigators may want to combine individual variables to create composites representing larger factors, and in this case, formal procedures of factor analysis should be followed. In the case of a new measure, in particular, exploratory (rather than confirmatory) factor analysis should be carried out to ensure that variables are being collapsed into factors that reliably represent the data. Although we do not have the space to elaborate on these procedures at length in this piece, we do recommend that users of observational data familiarize themselves with relevant literature and follow widely recommended procedures for factor analysis.

Our last, and perhaps most important, recommendation is that all analyses conducted on classroom observation data account for the multileveled nature of classroom data. We identify four potential levels that may need accounting for. First, if the same students were assessed across multiple time-points, variability within the same student between time-points will need to be accounted for. Second is variability between individual students, third is variability between classrooms, and fourth is variability between schools. Depending on the size and scope of your study and whether the design includes repeated observations, you may need to account for some or all of these levels. To determine whether enough variance exists at a given level to justify accounting for that level in analyses, we recommend calculating intra-class correlation coefficients (ICC; Dyer, Hanges, & Hall, 2005) and design effects (Peugh, 2010). ICCs provide estimates of the amounts of variance in each variable attributable to each level of the data, and design effects provide an indication of whether there is enough statistical power at a given level to justify accounting for that level in analyses. Together, an ICC of above .05 and a design effect greater than 2 indicate a need to account for that level in analyses. Many statistical computing programs are equipped with tools for multilevel analysis, including SPSS (IBM, 2011), MPlus (Muthén & Muthén, 2007), and HLM (Raudenbush & Bryk, 2002). In the study of focus, we conducted analyses using HLM because this analytic software provides especially robust estimations when using multileveled data. Although HLM is our personal favorite, other programs have notable strengths as well; for example, MPlus is especially tactful in its handling of missing data and its ability to estimate multileveled growth models. The choice of program is up to the investigator depending on their unique needs and existing resources; just know that this choice will need to be justified in future publication efforts.

Conclusion

The classroom is one of the most significant systems of developmental influence that children participate in (Bronfenbrenner & Morris, 2006), and as such, this setting deserves careful consideration in the field of educational research. Classroom observation research using video data allows for detailed investigation into many important classroom processes that likely play a large role in shaping how students develop. In addition, when collected and maintained properly, video observation data have the potential to support multiple investigations. As such, we view the methods discussed in this case study as extremely "high yield" in terms of their potential to contribute to the field. On the other side of this, though, is the fact that collecting, assessing, and analyzing classroom observation data take comparatively more effort from investigators than most other methodologies. As such, the recommendations in this case study are meant to provide guidance to those seeking to develop their own observational tools and/or conduct classroom video observation research. By knowing some of the common pitfalls of this methodology, as well as by having a plan of approach thoroughly mapped out before beginning a study, we hope that those attempting to conduct this type of research can get the most out of these efforts. In doing so, investigators can make important contributions that can advance the field's ability to identify what processes in the classroom help (or hinder) positive student development.

Acknowledgements

Funding for the studies presented here was provided by the National Institute of Child Health and Human Development, R01HD48539, R21HD062834, and P50 HD052120 and the U.S. Department of Education, Institute of Education Sciences, R305A130058, R305H04013, R305B070074, R305A130517/R305A160404, R305N160050 and R305F100027. The opinions expressed are ours and do not represent the views of the funding agencies.

Exercises and Discussion Questions

1. Imagine you want to create a novel observational tool to asses a previously unobserved aspect of the classroom. Write a one- to two-paragraph description of what you want to observe, how your tool will observe it, and which foundational theory (or theories) you will use to drive the creation of your tool.
2. Say you have created this tool and are preparing to use it in a study. Write a one- to two-paragraph description of how you will prepare your observation team for using this tool and how you will ensure they use it reliably throughout the duration of your study.
3. Explain the difference between a frequency observational tool and a duration observational tool. Find two examples of each in the education research literature.
4. Explain the difference between a global observational tool and a student-level observational tool. Find two examples of each in the education research literature.
5. A group of researchers is conducting a study investigating teacher–student interactions during classroom instruction. At four time-points across a single academic year, they collect video observations of 30 classrooms in four schools, all in the same school district. They later apply a student-level observational tool to these video data. How many levels exist in

these data that may need to be accounted for in analyses?

Further Reading

Kelcey, B., & **Carlisle, J. F.** (2013). Learning about teachers' literacy instruction from classroom observations. *Reading Research Quarterly*, 48, 301–317.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85–112.

Strong, M., **Gargani, J.**, & **Hacifazlioğlu, Ö**. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367–382.

Web Resources

www.noldus.com

isilearn.net

References

Bronfenbrenner, U., & **Morris, P. A.** (2006). The bioecological model of human development. In *Handbook of child psychology*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470147658.chpsy0114/abstract

Carlisle, J. F., **Kelcey, B.**, & **Berebitsky, D.** (2013). Teachers' support of students' vocabulary learning during literacy instruction in high poverty elementary schools. *American Educational Research Journal*, 50, 1360–1391. doi:http://dx.doi.org/10.3102/0002831213492844

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213.

Connor, C. M., **Morrison, F. J.**, **Fishman, B.**, **Crowe, E. C.**, **Al Otaiba, S.**, & **Schatschneider, C.** (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, 24, 1408–1419.

Connor, C. M., **Morrison, F. J.**, **Fishman, B.**, **Ponitz, C. C.**, **Glasney, S.**, **Underwood, P.**, … **Christopher, E. C. C.** (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, 38, 85–99. doi:http://dx.doi.org/10.3102/0013189X09332373

Connor, C. M., **Spencer, M.**, **Day, S. L.**, **Giuliani, S.**, **Ingebrand, S. W.**, **McLean, L.**, & **Morrison, F. J.** (2014). Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes. *Journal of Educational Psychology*, 106, 762–778.

Dyer, N. G., **Hanges P. J.**, & **Hall R. J.** (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, 16, 149–167.

IBM. (2011). *IBM SPSS statistics for windows, version 22.0*. Retrieved from https://www-01.ibm.com/support/docview.wss?uid=swg21646821

Kane, T. J., & **Staiger, D. O.** (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (Research paper, MET Project). Seattle, WA: Bill & Melinda Gates Foundation.

McLean, L., & **Connor, C.M.** (in press). Relations between third-grade teachers' depressive symptoms and their feedback to students, with implications for student mathematics achievement. *School Psychology Quarterly*.

McLean, L., **Sparapani, N. E.**, **Toste, J.**, & **Connor, C. M.** (2016). Classroom quality as a predictor of first grader's time in non-instructional activities and literacy achievement. *Journal of School Psychology*, 56, 45–58.

Muthén, L. K., & **Muthén, B. O.** (2007). *Statistical analysis with latent variables* (Version 3). Los Angeles, CA: Muthén & Muthén.

Nielsen, J. (2007, August). Ten usability heuristics mean for MIMAS websites and web based interfaces. *Design*. Retrieved from http://web-standards.mimas.ac.uk/resources/web-guidelines/topics/appendix1/usability/10heuristics-mimas.pdf

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85–112.

Raudenbush, S. W., & **Bryk, A. S.** (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: SAGE.

Strong, M., **Gargani, J.**, & **Hacifazlioğlu, Ö** (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. *Journal of Teacher Education*, 62, 367–382.

Thelen, E., & **Smith, L. B.** (1996). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press.

Yoshikawa, H., & **Hsueh, J.** (2001). Child development and public policy: Toward a dynamic systems perspective. *Child Development*, 72, 1887–1903.