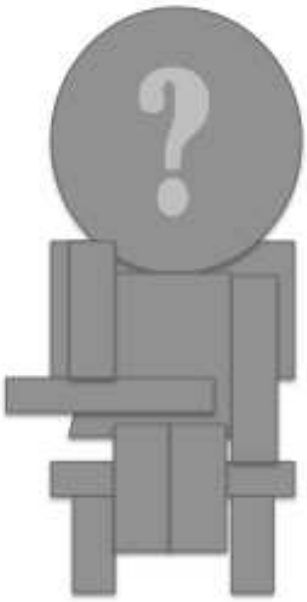


BROMP

The Baker Rodrigo Ocumpaugh Monitoring Protocol

# Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual

**Jaclyn Ocumpaugh,** Teachers College, Columbia University
**Ryan S. Baker,** Teachers College, Columbia University
**Ma. Mercedes T. Rodrigo,** Ateneo de Manila University

## Acknowledgements

This report was prepared in part with funding from the National Science Foundation through the Pittsburgh Science of Learning Center, Award # SBE-0836012, and the Bill and Melinda Gates Foundation, Award #OPP1048577. We would also like to thank the colleagues who have refined the method by learning and using it.

## TABLE OF CONTENTS

# Chapter 1: History and Purpose

## 1.1 Introduction

The **Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0** is a method for Quantitative Field Observations (QFOs) of student behaviors and affective states in classroom environments. Formerly known as the Baker Rodrigo Observation Method Protocol, BROMP is an *momentary time sampling method* in which trained, certified observers record students' behavior and affect individually in a pre-determined order. Although developed for observations of students using educational software, BROMP is easily adaptable to other environments where student engagement is of interest. Today, BROMP is implemented by certified coders using the *Human Affect Recording Tool* (HART), which offers a variety of behavior and affect coding schemes that are relevant to understanding the relationships between students' engagement and their learning environments.

BROMP coding was first used in 2004 to record educationally relevant behaviors during Baker's early research on *gaming the system* (Baker, Corbett, Koedinger, & Wagner, 2004). In this study, students were coded as on task but working alone, on task but participating in conversation, off-task or gaming the system, following conventions similar to previous research in this area (cf. Karweit & Slavin, 1982; Lloyd & Loper, 1986). In 2007, during work with Rodrigo, a coding scheme for affective states was added to the protocol (Rodrigo et al., 2007), based on evidence from Graesser and his colleagues as to which affective states were frequent during real-world learning and associated with differences in student outcomes (D'Mello, Graesser, & Picard, 2007). The protocol was described more thoroughly in Baker, D'Mello, Rodrigo, and Graesser (2010) and formalized in 2012 with the publication of the first training manual (Ocumpaugh, Baker, & Rodrigo 2012).

BROMP observations are often used to obtain *ground truth* labels for *Educational Data Mining* (EDM) research. In these studies, BROMP codes are synchronized to log-file data compiled by the educational software that students are using at the time of the observations. EDM researchers then use data mining algorithms to determine what patterns in the software interactions correlate with the field observations. This process (described in detail by Baker et al., 2012, Pardos et al., 2013, and other articles) results in automated models, called *detectors,* that infer when students are *bored, frustrated, confused, off-task*, etc. These detectors were originally created with a goal of supporting automated intervention and teacher reporting. Although used for these purposes (Baker et al., 2006; Ocumpaugh et al., in preparation), these detectors have thus far achieved greater usage as components in *discovery with models* research, where a data-mined model of one construct is used as a component in the analysis of another construct (Baker & Yacef, 2009).

BROMP assumes that behavior and affect are at least partially orthogonal, and so they are coded simultaneously but separately (e.g., a student could be *gaming the system* and *bored*, or *gaming the system* and *frustrated*). Coding schemes that have been developed for a variety of learning environments are available in HART (the app used to implement BROMP), and it is possible to create new coding schemes. This allows researchers to include categories that are unique to (or more common within) a particular learning environment. (For instance, *surprise* and *delight* are rare in many educational settings, but they are often common in software environments with a

game-like design.) BROMP-certified coders are trained to identify behaviors or affective states that do not match their current coding scheme, classifying them as "other" using the "?"code.

BROMP-certified observers base their judgment of a student's affective state or behavior on the student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students. This contextualized coding practice is in line with Planalp et al.'s (1996) descriptive research on how humans generally identify affect, using multiple cues in concert for maximum accuracy rather than attempting to select individual cues. In our experience, attempting to focus on specific cues (rather than using more holistic judgments of behavior and affect) substantially reduces inter-rater agreement.

## 1.2 Previous Research Using BROMP

Previous research has produced a number of successful EDM models based on BROMP observations. In terms of behavior, these methods have been used to develop and validate models that can infer gaming the system (Baker, Corbett, & Koedinger, 2004; Baker, Corbett, Roll, & Koedinger, 2008) and off-task behavior (Baker, 2007). In terms of affect, these methods have been used to develop and validate models that can infer confusion, boredom, frustration, and engaged concentration (Baker et al., 2012, Pardos et al., 2013). These methods have also been used to study affect and the contexts and fashions in which affect emerges and changes over time (see Rodrigo et al., 2007, 2008a, 2008b; Baker, Rodrigo, & Xolocotzin, 2007; Baker, D'Mello, Rodrigo, & Graesser, 2010; Baker, Moore, et al., 2011; San Pedro et al., 2011; Hershkovitz et al., 2012; Liu et al., 2013; Rodrigo, Baker, & Rossi, 2013).

BROMP has been used to observe students from kindergarten to college. It is well established as a method for studying how affective and behavioral constructs manifest in computer-based learning environments (see Table 1). Initially, these models were just for behavior, but in 2012, the BROMP was used to create the first cross-validated, interaction-based models of affective states for Cognitive Tutor Algebra (Baker et al., 2012).

Increasingly BROMP is also being used to study learning environments that do not include technology (e.g., Godwin et al., 2014 and Fisher et al., 2014). It has also been used to estimate engagement in technology systems without first constructing EDM models (e.g., Ocumpaugh, Baker, Gaudino, Labrum, & Dezendorf, 2013). In these studies, researchers typically calculate the percentage of time that students were observed in each affective or behavioral state in the coding scheme. This new use of BROMP can provide researchers, educators, and policy makers with important information about student engagement, however, we urge researchers who are planning to use BROMP in this way to consult carefully with BROMP developers to ensure that sampling conditions meet validity requirements.[1]

[1] Note that Pacquette, Ocumpaugh, & Baker, (in preparation) have developed a computerized simulation program for testing the validity of interval time sampling methods given specific parameters. Researchers may also be interested in *ARPObservations,* a freely available tool for testing momentary time sampling in the R platform (Pustejovsky & Runyon, 2014).

[2] It is rare for individuals to be successfully BROMP-certified for a culture they are not native to. Currently three individuals have been BROMP-certified out of their native country.

[3] There are currently 40 coding schema available in the HART application used to make BROMP recordings (see Appendix B), and it is possible to code new ones (see Section 4.2).

**Table 1: Educational software that has been studied using BROMP**

| System | Developer | Subject Matter |
|---|---|---|
| Aplusix | Laboratoire d'Informatique de Grenoble | Arithmetic & Algebra |
| ASSISTments | Worcester Polytechnic Institute | Mathematics |
| BlueJ | University of Kent | Computer Programming (Java) |
| EcoMUVE | Harvard | Environmental Science |
| Ecolab/M-Ecolab | | Ecology |
| Chemistry Virtual Laboratory | Carnegie Mellon University | Chemistry |
| Cognitive Tutor Algebra | Carnegie Learning/Appollo | Algebra |
| Cognitive Tutor Geometry | Carnegie Learning/Appollo | Geometry |
| Cognitive Tutor Middle School | Carnegie Learning/Appollo | Mathematics |
| The Incredible Machine | Jeff Tunnell & Chris Cole | Physics |
| InqITS (formerly Science ASSISTments) | Worcester Polytechnic Institute | Science Inquiry Skills |
| Middle School Mathematics Tutor (Scooter the Tutor) | Carnegie Mellon University | Mathematics |
| Physics Playground (formerly Newton's Playground) | Florida State University | Physics |
| Reasoning Mind | Reasoning Mind | Mathematics |
| Refractions | University of Washington & Utah State University | Fractions |
| vMedic | U.S. Army | Army Field Medicine |

Although most of the research using BROMP has taken place within the United States, where Ocumpaugh and Baker direct training, the formalization of the method included the training of BROMP-certified coders in the Philippines, work led by Rodrigo. In 2014, coding schemes were developed and tested under the direction of Viola Krishnamani and Chokanath Hymavathy in India, who are rapidly expanding the program there (Hymavathy, Krishnamani, & Sumathi, 2014). As of this writing, there are nearly 130 BROMP-certified coders worldwide, with more in India than the United States and Philippines combined, and efforts are currently underway to adapt BROMP for use in Mexico.

Because of legitimate concerns about cross-cultural differences in the presentation of affect (e.g., Elfenhein & Ambady, 2003; Jack, Garrod, Yu, Caldara, & Schyns, 2012; although see Sauter & Eisner, 2013) and the linguistic labeling systems used to identify affect (e.g., Lindquist & Gendron, 2013), the procedure for adapting BROMP to a new country, including the training of the first coders, is a collaborative process involving both experienced members of the BROMP development team and researchers who are native to that culture. Once BROMP-certification is complete for the first (native) researchers, researchers from other backgrounds may attempt to

become BROMP certified for that country.[2] These efforts have facilitated cross-cultural comparison studies of engagement (e.g. Rodrigo, Baker, & Rossi, 2013), and we hope that more such studies will follow as BROMP is used more frequently, improving the diversity of populations studied within educational research communities (see Blanchard 2012, 2014).

**Figure 1: Countries that have BROMP-certified coders to date.**

## 1.3 Development of BROMP Coding Schemes for Indicators of Student Engagement

In this section, we provide a very brief discussion the relevance of behavioral and affective indicators in understanding student engagement as well as an overview of the affective and behavioral categories that are typically used in BROMP coding schemes. (A fully defined list is available in Appendix A). In particular, we concentrate on the design of the schemes that are most commonly used in BROMP research—those developed for the Pittsburgh Science of Learning Center's (PSLC).

Behavioral coding schemes for direct observation are not new in education research (see Bakemen, 2000; Volpe et al., 2005), and it is becoming increasingly common to see coding schemes that look at other issues related to engagement and cognition (see for example, reviews in Boekarts, 2007; Linnenbrink-Garcia et al., 2011a, 2011b). BROMP was one of the first direct observation schedules designed specifically for *in situ* observations of technological classrooms, and it has developed over time to include observations of behavior (e.g., *on task, on-task conversation, off-task, gaming the system,* and *?*) and educationally relevant affective states (e.g.,

[2] It is rare for individuals to be successfully BROMP-certified for a culture they are not native to. Currently three individuals have been BROMP-certified out of their native country.

*boredom, confusion, engaged concentration, delight, frustration,* and *?*). Less commonly, BROMP is also used in conjunction with observations of teacher behaviors (e.g., Godwin et al., 2014). As with the development of other classroom coding schemes, the inclusion and exclusion of constructs has depended upon the research question. In initial BROMP studies, which were designed to train software how to recognize student engagement indicators, teacher behaviors were not necessary, nor were certain specific differentiations (e.g., a student who was off task and reading vs. a student who was off task and out of her seat). As BROMP has evolved to address different research questions new coding schemes[3] have been developed to address those needs.

The first BROMP behavioral coding scheme was developed for studies of educational software that were being conducted for the Pittsburgh Science of Learning Center (PSLC). The initial goal of this research was to study student learning within the software, but students weren't making the predicted gains. Fieldwork revealed that many students were gaming the system, a behavioral pattern that is neither strictly on task or strictly off task, since students were using the learning software but exploiting the properties of the system that let them advance without actually learning the material. Subsequent research on this behavior in several systems has shown that this behavior is linked to poorer learning outcomes (Baker, D'Mello, Rodrigo, & Graesser, 2010; Beck & Rodrigo, 2014; Cocca, Hershkovitz, & Baker, 2009; Pardos et al., 2014; San Pedro, Baker, Bowers, & Heffernan, 2013). As a result of research using BROMP behavioral coding schemes, several learning software systems[4] now have stealth detectors that are triggered when student interactions with the software suggest that they are gaming the system or off task.

The addition of affective coding schemes has been instrumental in ongoing research to improve student-learning outcomes, but it often raises questions among those who are first learning about BROMP, especially with the increased use of sensors and survey data to assess these same constructs. The idea that BROMP coders are assessing affect with observations alone sometimes raises concerns among people who are not familiar with the research literature on affective states. These are valid concerns, since self-presentation effects are a real possibility when students feel like they are being watched (CITATION). While we cannot control for the effects that other students or the teacher might have on this presentation, observer effects can be substantially minimized in classroom environments, and coders are trained to minimize their presence in the classroom during the certification process (see Section 3.4). It is also worth noting that self-presentation effects may be a concern in other methods (e.g., surveys or video), and that studies with more sophisticated sensors, including posture sensors, heart rate and sweat monitors, and even MRIs, rely on either observation or surveys to interpret the patterns in their data. While survey data has the benefit of coming directly from the student, there are other concerns beyond the self-presentation issues (see discussion in Baker & Ocumpaugh, 2015), including the self-awareness (e.g., Bieg et al., 2014) and intersubject reliability issues (e.g., Porayska et al., 2013). That is, we would like to acknowledge the legitimate concerns that people might have about using observational methods in affective research, but we feel it is also

[3] There are currently 40 coding schema available in the HART application used to make BROMP recordings (see Appendix B), and it is possible to code new ones (see Section 4.2).
[4] These include ASSISTments and Cognitive Tutor.

important to point out that affect is inherently difficult to define, and that other methods often rely on observation to validate their findings[5].

BROMP coding schemes for affective engagement indicators were developed based on previous research on relevant emotional states (e.g., D'Mello, Graesser, & Picard, 2007). Ekman, one of the most influential researchers in the research of emotions, identifies several basic emotions that are described as culturally universal (Ekman & Frisen, 1971), but other researchers have raised concerns about this work (see discussion in Kory & D'Mello 2105). More importantly, many of these emotions are rare in educational settings, and those that are common do not appear to be strongly correlated to learning (Lehman et al., 2008). BROMP coding schemes have been tailored instead to focus on affect that is relatively common in educational settings and associated with differences in learning outcomes. The PSLC coding scheme, for example, includes *boredom*, which is linked to poorer learning outcomes (e.g., Csikszentmihali, 1990; Miserandino, 1996) and *engaged concentration*, which is linked to improved learning outcomes (e.g., Csikszentmihalyi, 1990), but also *confusion* and *frustration*, which have mixed effects on learning (e.g., Liu et al., 2013; Kort et al., 2001; Patrick et al., 1993; Graesser & Olde, 2003; Kort et al., 2001).

BROMP coding schemes are also designed to address methodological concerns. When coding schemes are designed to be used as training labels for EDM detector development, it is also important to choose categories that can reasonably by inferred from the log files of a student's interactions with the software being studied. That is, a construct like *off-task* could likely be predicted from the student's interactions with the software (and it has, see Baker, 2007 and Cetintas et al., 2010), but more specific types of off-task behavior (e.g., whether the student is talking to a neighbor or reading a magazine) are not something that could be predicted from a log file, and so that level of detail is not recorded. When researchers work in other domains, greater specificity may become relevant (e.g., Godwin et al., 2013 looks at the type of off-task behavior in completely offline research), but researchers should consult with BROMP developers before adding new constructs, especially those which occur infrequently, since some research suggests that rare constructs may be problematic for both inter-rater agreement and for prevalence estimates in momentary time sampling.

The addition of coding schemes that describe the behavior of the teacher or the activities of the class has been quite recent. In large part, this has grown out of the expansion of BROMP to look at student engagement in non-technological classrooms. For example, Fisher and her colleagues at Carnegie Mellon University have used a second observer to code these conditions in kindergarten classrooms so that engagement can be monitored as classroom activities change (e.g. Godwin et al., 2013, 2014; Fisher et al., 2014). However, even in technological classrooms, additional information may be useful. Our partners at Reasoning Mind, which provides a blended learning system (merging software with teacher training), have used these practices to examine the extent to which their teacher-training efforts have helped.

[5] Please see additional discussions related to these issues Chapter 2.

# Chapter 2: Training and Inter-Rater Agreement

## 2.1 BROMP Training Process:

We have a relatively short field training process for BROMP that ensures that field observers are coding accurately and consistently. In addition to the information presented in this manual, we provide additional training during the certification process. This section provides an overview of the process we use to certify new trainers. In general, there are three phases to this process: (1) pre-field training, (2) field training, and (3) inter-rater agreement testing. Typically pre-field training takes places a few days before entering the field. Once we enter the field, the amount of time taken to successfully become certified varies based upon a number of factors, including the schedule of the school where the training is taking place. Although many people can complete this process in a single day, this assumes ideal conditions. It is not unusual for a novice to require 2 days in field to complete the process, and in some rare cases, 3 days may be required.

### 2.1.1 Phase 1: Pre-Field Training

In general, novices are asked to read the BROMP training manual (this document) before beginning the initial training session with a certified coder. In the initial training session, it may be worthwhile for several novice coders to meet with the trainer at the same time to facilitate greater discussion. During this phase of the process, novice coders are introduced to the general concepts that the field observations are designed to address. (See Chapter1.) They are also given an opportunity to interact with the HART, the android application that we use to record data. Short videos are sometimes played,[6] and discussions of the coding scheme, including ambiguous examples, take place so that novice coders have the opportunity to ask questions about each of the constructs they will be observing.

### 2.1.2 Phase 2: Field Training

Once we enter the field, a novice coder shadows an experienced, certified trainer. The trainer helps the novice practice entering information into HART while quietly discussing the coding process for each student being observed. In this way, the novice has further opportunities to familiarize himself/herself with HART and an opportunity to ask more questions about the labeling scheme before checking IRA.

The goal of this phase of the training process is to ensure that the novice is sufficiently comfortable with both the technology and the coding process before the testing for IRA. The amount of time for this varies, but it can often be accomplished in 2-3 class periods. It is not advisable for more than one novice to participate in this process at the same time as it may substantially increase students' sensitivity to the observers.

[6] We do not make BROMP training videos publically available, but some people find it useful to spend some time watching naturalistic videos of classroom time. See Appendix C for links to classroom observations that are publically available online.

### 2.1.3 Phase 3: Inter-Rater Agreement (IRA) Testing:

During the IRA test, we check to see that the novice understands the process but *AND* is coding consistently with the expert trainer. (That is, we use a *criterion-referenced agreement* test, e.g., Subkoviak & Baker, 1977) In sessions where we test this, the novice and the trainer will work together to signal which student is being coded and when. Since the goal is to code the first affect/behavior witnessed, this requires some sort of hand signal or countdown so that both coders are beginning their observation at precisely the same time. This is particularly important in classrooms where students are cycling quickly from one behavior or affective state to another. If the novice does not pay attention to the trainer's cues on when to start the observation of each student, they are unlikely to demonstrate IRA.

We typically recommend having at least 60 observations before we check for consistency between the novice and the trainer. Some research also suggests that the trainer should have observed at least 10 instances of each construct before IRA is checked (see Cicchetti, 1994; Watkins & Pacheco, 2000.) While this is a good guideline, particularly for researchers who wish to study the prevalence of a given construct from BROMP field research, it is sometimes difficult to perfectly reach this goal.

If acceptable inter-rater agreement is not achieved, the trainer will analyze which situations the trainee disagreed with them in, and discuss these situations. Typically, additional field training focusing on areas of disagreement is conducted before attempting to test for inter-rater agreement again. Many novice coders must have their inter-rater r agreement checked several times before they can be certified. If a novice has not achieved IRA after 3 rounds of testing, the trainer will make a judgment call about whether or not it is productive to continue.

## 2.2 Inter-Rater Agreement (IRA) and Validity:

BROMP inter-rater agreement (IRA) is calculated using Cohen's (1960) Kappa, which scales from -1 to +1, capturing how likely it is that agreement is due to chance. Kappa is recognized across disciplines as a well-established metric for IRA (See for example, Fleiss, 1981; Perreault & Leigh, 1989; Tooth & Ottenbacher, 2004; Wirtz & Kutschmann, 2007; Sadatsafavi, Najafzadeh, Lynd, & Marra, 2008; Dewey, 1983.) Kappa is preferred over accuracy, a metric that does not account for chance agreement.

$$\mathrm{K} = \frac{Observed\ Agreement - Chance\ Agreement}{1 - Chance\ Agreement} = \frac{P_o - P_c}{1 - P_c}$$

For perfect performance, Kappa = +1. Kappa > 0 represents performance above chance; Kappa < 0 represents performance below chance. However, Kappa itself does not reveal whether coding differences are random, caused by synchronization failure, or caused by systematic differences between coders; this determination requires additional qualitative examinations of the data (Sim & Wright, 2005).

Cohen's (1968) refinement of Kappa allows for the seriousness of different measurement errors to be weighted during the correlation, but no such assumptions are made when calculating IRA for BROMP. Instead, Cohen's (1960) original unweighted Kappa is calculated separately for behavior and for affect. During this calculation, any observations (whether made by the trainer or the trainee) that include a "?" are discarded from the calculations. That is, if one of the coders records a '?' when the other records confusion, that observation is discarded before calculating Kappa for the affective coding schemes. However, the observation is only discarded in the coding scheme where it was labeled as "?". If both coders were able to code behavior (neither used a '?' for the behavior code) during an observation that received a '?" for affect, then the behavior codes are still used to calculate IRA.

A number of different researchers have made recommendations about cut-offs, ranging from a Kappa of 0.4 to 0.8, for two raters to be considered to have acceptable agreement (Landis & Koch, 1977; Fleiss, 1971; Di Eugenio & Glass 2004; etc.). These numbers are inherently arbitrary, and in fact Kappa does not have invariant meaning across data distributions (e.g. 0.8 is harder in one data set than another). Higher expected numbers are typically seen in domains where coders are working using definitions that are not open to judgment. Lower expected numbers are typically seen in domains where the underlying truth is uncertain, where coding schemes are holistic, and where data coding is cheap (as a large amount of imperfect data can be more reliable in aggregate than a small amount of excellent data). In practice, we use a cutoff of 0.6 for BROMP certification, where the coder must achieve 0.6 or higher for both schemes. Typically we find that Kappa is higher for behavior coding schemes than for affect coding schemes.

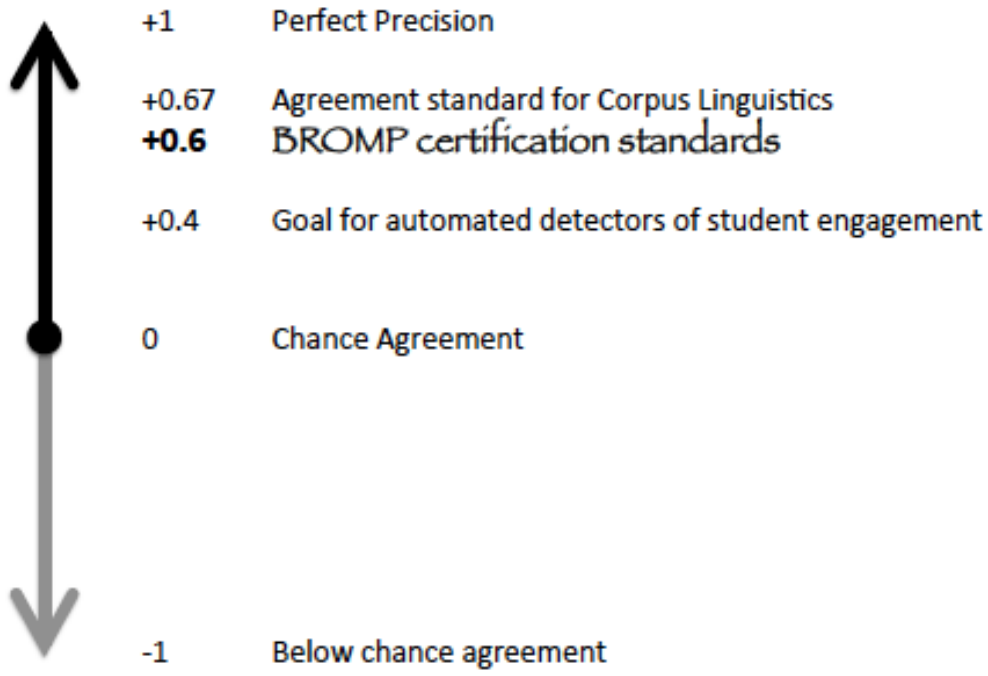



**Figure 2: Interpreting Kappa in the Context of BROMP**

Achieving IRA with a certified coder is especially important when coding constructs like affective states, where a gold-standard is impossible to obtain. It is quite possible that self-report surveys, peer-evaluations, teacher evaluations and BROMP observations could all result in different affective states being reported for a given student at a given time (see Porayska-Pomsta et al., 2013 for a discussion of this), and it is not clear that any one of those data sources would be more true than another (Bieg et al., 2014). Direct, *in situ* observation schemes like BROMP may not always agree with other evaluation methods, but for constructs like affect, where it is possible to experience more than one emotion at a time, the fact that a student and observer might come to different conclusions about which affective state is most relevant does not necessarily negate the validity of either judgment.

This problem has been extensively discussed in other literature (e.g., Poranyska-Pomsta et al., 2013; Afzal & Robinson, 2011; Pantic & Rothkrantz, 2003). Although popular belief (and the occasional journal reviewer) may lean towards treating some data as more valid than others, all require normalization. For example, self-report studies and sensors, which might be argued to be the most direct measures of a student's affective state, have their own challenges. Frequent self-report surveys interrupt student learning in ways that interfere with the measurement process, while retrospective reports may be susceptible to reinterpretation and other memory-related processes. In fact, the very act of labeling an emotion has been shown to alter its physiological response patterns (Kassam & Mendes, 2013). What's more, self-report techniques may not work well in cases with students who are not especially self-aware. This is of particular concern when researching young children, who are reasonably good at identifying valance, but have not yet developed adult-like categories for emotion (see Widen & Russel, 2008). Adults are usually more skilled at recognizing emotions, but remain susceptible to self-presentation and demand effects. Sensors may be able to address some of these concerns, in that they operate in real time and are less likely to cause regular interruptions. However, sensor accuracy is still imperfect and sensors can be costly and are sometimes still invasive (e.g., requiring students to adjust web cams or wear skin conductance sensors). Breakage rates for many sensors remain high in real-world classrooms and even in laboratory settings. And all sensor-based models are originally based either on self-report, video coding, or field observations. .

BROMP coders are trained to identify the constructs that are most prominent while taking into account the full classroom context, meaning they should be attuned to individual variation. However, it is still important that these standards be applied consistently. For this reason, achieving an adequate inter-rater agreement (Kappa > .6) with a certified coder is a requirement for anybody reporting to use BROMP.

At this time, obtaining a sufficient Kappa on a single coding scheme is sufficient for BROMP certification. However, you may want to consider re-checking inter-rater agreement if you are working with a radically different coding scheme than the one that you were trained upon or if you are creating a new coding scheme for a specific learning environment. The authors of this coding manual are happy to advise you in doing so. It is important to ensure that any new categories are being coded consistently. For constructs that are quite easily defined, (e.g. *creative metanarrative*, Ocumpaugh et. al., 2014), it may be possible to simply add the new category to your coding scheme. For more ambiguous constructs, it may be advisable to replicate the inter-rater agreement process with another member of your research team to ensure that it is being

consistently identified and coded.[7] Please see our Reporting Standards (Chapter 5) to ensure that the addition of any new construct be thoroughly documented and reported.

Occasionally, journal reviewers will make recommendations about the BROMP certification process, including suggestions that IRA be re-checked every time a new learning environment is observed. While we have no objections to BROMP observers resynchronizing with one another, it is not always practical (or even possible) to conduct such normalization procedures at every field site. In some cases, as when coding an event of a limited duration, such requirements would prohibit research from occurring. Furthermore, some research suggests that over-practice could be counterproductive, particularly when coding schemes are being applied to contexts and constructs where there are inherent ambiguities. (See, for instance, Towstopiat's (1984) discussion of Medley and Norton (1971), where they suggest that "brainwashing" observers into consistency when coding behaviors that are inherently ambiguous introduces bias towards one category, whereas some degree of disagreement on ambiguous behaviors more accurately represents real-world conditions.)

## 2.3 Cross-Cultural Coding and Reliability:

There is now an extensive body of research on the cross-cultural identification of emotions which we invite readers to familiarize themselves with (e.g., Baron-Cohen et al., 2004). A number of issues exist when it comes to making recommendations on the best-practices for dealing with *intercultural* (across culture) and *intracultural* (within culture) social differences, including the background of the subjects, the background of the observers, and the types of data being examined. Strong evidence suggests that when presented with static photographs of people from a different background than the observer, the categorization of emotions is less reliable (Krumhuber, Kappas, & Manstead, 2013), and further evidence suggests that different cultures may be more reliant on context than others (Matsumoto, Hwang, & Yamada, 2010). However, there are also concerns about what standards are being used to validate these codes. Judgments of correctness could be based on self-reports, on peer judgments, or on research-based coding schemes like Ekman's FACS, and results will vary accordingly. What's more, there is evidence that cross-cultural identification weaknesses might be mitigated if the observer has had more contact with the group in question (Beaupré and Hess, 2006) or when the observer is presented with more dynamic information (see review in Krumhuber, Kappas, & Manstead, 2013).

BROMP has been developed with these issues in mind. Emotional categories selected for BROMP coding schemes are those thought to be most educationally relevant (D'Mello, Graesser, & Picard, 2007), with adaptions made to address culturally-specific differences in the appropriateness of different constructs. For example, when BROMP was adapted to India, we worked with local educational researchers to adjust to local norms. In this case, we learned that *frustration* was considered inappropriate to express (and would therefore be rare); therefore,

[7] Of course, there are times when it is impossible replicate the research conditions that have induced a new construct. For example, it might be possible to induce the affective state of boredom by giving students repetitive, tedious tasks, which could lead to students going off task. However, it could be more difficult to induce *creative metanarrative*, a highly imaginative behavior identified by Ocumpaugh et al., (2014), where students created an alternative storyline for the virtual environment. (For example, one student repeatedly discussed interactions his avatar had with characters that were not in the game, including police officers and prostitutes.)

BROMP-certified observers in India are trained to code for *contempt* instead.

As discussed above, there are a number of concerns about in-group advantage for identifying emotion (e.g., Elfenbein & Ambady, 2002a, 2002b, 2003). For this reason, we recommend that observers be of a similar background to the students they are observing. This is particularly important during the early stages of observations of a new population, and for this reason we have ensured that only observers who are native to the country being observed were used to establish initial measurements of inter-rater agreement for that country. Once the coding scheme is well established, it is sometimes possible to certify coders from other populations, but in our experience non-natives sometimes have difficulty passing inter-rater agreement checks for affect coding, even when restricted to educationally relevant categories.

# Chapter 3: Useful Tips for New BROMP Observers

## 3.1 Overview of Chapter 3

This chapter provides bulleted discussions of things that often raise concerns for new BROMP observers. These include best strategies for conducting unobtrusive observations, a short overview of the typical coding schemes, and discussions of how to deal with ambiguity in the coding schemes.

In general, it is a good idea for new trainees to familiarize themselves with this chapter before the training process, but we do NOT recommend trying to memorize it. Your trainer will go over these things again with you as part of the process, and we feel that attempts to memorize these things will only serve to make new coders nervous. (In our experience being in distress typically makes people less reliable coders!)

We also encourage previously certified coders to re-familiarize themselves with these guidelines in an effort to minimize coding *drift.* This is particularly important for coders who have been out of a field for quite a long time.

## 3.2 Typical Coding Schemes:

BROMP *typically* uses a dual coding scheme, recording behavior simultaneously, but separately from affect. There are several coding scheme choices to choose from when you start a session, and it is possible for a programmer to customize a new coding scheme that can be installed on the phone (see Section 4.2). It is also possible to use a third coding scheme for interventions, classroom activities, or other activities related to the teacher.

Once you have selected these coding schemes, HART will automatically present you with drop-down menus that include ONLY the constructs that were already programmed for those coding schemes. An expanded list of coding schemes is included in Appendix B and descriptions of each construct are included in Appendix A, but for many learning environments, the PSLC behavior and affect coding schemes (Shown in Table 2) are preferred.

**Table 2: Coding Schemes developed for the PSLC**

| Behavior | Affect |
|---|---|
| On task | Boredom |
| On-task Conversation | Confusion |
| Off-task | Engaged Concentration |
| $ = gaming the system | Frustration |
| ? = other | ? = other |

Note that "?", "?" is used when a student could not be coded for either behavior or affect. In such cases, it is possible that the student was displaying behavior that was not clearly on-task or off-task, and his or her affective state was not part of the coding scheme. However, it is also possible that the student:

- was out of his or her seat (although it is often still possible to code such students)
- was out of the room
- became sensitized to the field worker

These codes are typically excluded from EDM models (since it is hard for software to predict when a student needs to go to the bathroom, for example), but they may be important for researchers who are conducting other kinds of studies. For example, if an observer is unable to code a large number of students who are out of the classroom due to behavioral issues, researchers who are using raw BROMP observations (as opposed to data mining models) to study the frequency of behavioral constructs may need to modify the protocol to account for this data. Future iterations of HART may include a missing code or skip button to help differentiate these contexts.

## 3.3 Notes on ambiguous behavior or affect:

You will undoubtedly encounter numerous instances during your observations where a student may be doing more than one thing at a time. If the student is engaging in one behavior (or affect) and then switches to another, code the first behavior (or affect) seen, as that is less ambiguous. If two things are occurring simultaneously, you should use your best judgment to determine the predominant behavior/affect that the student is presenting OR you should code a "?" (the catch-all category for anything that doesn't fit our coding scheme). Behavior and affect are cultural constructs that we learn to identify as we grow up in that culture. We use clues like facial expressions, body language, vocal expression, and other contextual clues to make holistic decisions about what other people are doing and thinking all day. Most people are competent at doing this, but it is important that BROMP labeling is consistent. This section is meant to give you some guidelines for common questions that new coders have.

### 3.3.1 Examples of Ambiguous Behavior:

1. A student is waiting with their hand in the air. They aren't working on their assignment within the software, but that appears to be because they are legitimately waiting for the teacher. This should probably be coded as *on-task*, even though they are not engaged with the software at the time. Of course, you should use your best judgment. If a student seems to be manipulating the teacher to avoid work, you should code that student as *off-task*.
2. Similarly, if a student is doing scratch work on paper, they may not appear to be engaged with the software. However, if they are using that scratch work in a manner that is consistent with the learning task, you should still code them as *on-task*.
3. If you may see two students who are working but sporadically discussing something not related to the assignment (say, last night's television show), you should use your best judgment to determine which behavior is predominant. Remember to consider things like posture, body language, and facial presentations (e.g., are they mainly looking at their own computer screens, or is there a lot of turning to face each other during the conversation). If

the student you are coding appears to be mostly concentrating on the educational task, you can code that student as *on task conversation*. If, however, that student is more oriented towards their classmate, you should probably code them as *off-task*.

4. A student first appears to be *off-task* because he or she is staring into space, but then you notice that he or she is responding to a teacher or another classmate who is asking them an *on-task* question. You should consider coding the student as *on-task conversation* if their responses indicate that they were paying attention the whole time.
5. A student is using an online search engine (e.g., Google). If it looks like the student is searching for course-related information (and the teacher has not forbidden this task, e.g., in the cases of a testing situation), then it is probably okay to code *on task*. However, if the student is checking local movie listings, *off-task* is a more appropriate code.
6. In some cases, it's not clear whether interaction with an online search engine (as in #5, above) is strictly *off-task* or *on-task*. For example, if it looks like a student has finished his or her assignments and is checking the local movie listings while talking a peer through their questions and waiting for further instruction from the teacher, *on-task* might be the most appropriate code. If a pattern like this is quite common, every effort should be made to label these instances and the researcher should thoroughly document and report his or her decision making process. (This is particularly important when BROMP observations labels are not being used as the ground truth to train EDM models.) If such an example is a relatively rare occurrence, however, a '?' may be more appropriate.
7. When in doubt, always use the "?" option.

### 3.3.2 Discussion of Ambiguous Affect

Affect is somewhat trickier to code since you are more likely to see overlap in affective states than you are with the behavioral constructs that we code for. For example, you are more likely to see a student who is confused become frustrated than you are to see a student who is both on- and off- task at the same time (although note #3 above!). In cases such as these, if one affective state clearly precedes the next, choose that one. Otherwise, you should choose the affect that appears most prominent. If a student appears more focused than *confused*, code *engaged concentration*. If a student appears *confused* and *frustrated*, and the frustration is strong, code *frustration*. However, if you're not sure what the student's affect is, use the "*?*" option.

Many novice coders struggle because they lack confidence in making these distinctions. Sometimes the presentation of an affective state is quite obvious, but other times affect categories show considerable overlap in facial expressions and posture. Indeed, a student who is experiencing *engaged concentration* may have a facial expression (a "scowl") that is quite difficult to distinguish between *confusion* and/or *frustration* at first glance. Likewise, people who are *frustrated* often smile, but it is not a smile of genuine happiness. (That is, it is not the Duchenne smile.) If you are not sure, take a few extra seconds to make your decision. You should still code the first affect that you see; if their initial affect is uncertain but then clearly changes (e.g. you're not certain if they are *confused* or *bored*, but then an event happens across the room which *delights* them), you should use the "?" code. However, given a couple of extra seconds you might see sighs, fist pounding, or just general persistence that will help you to determine which affect the student is displaying.

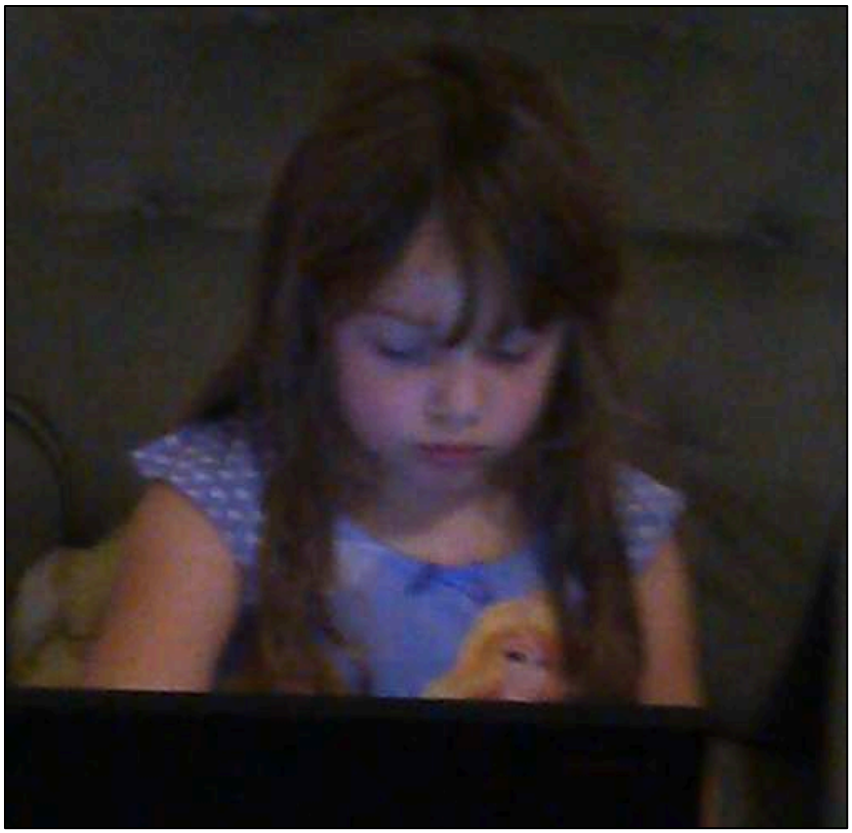


**Figure 3: Engaged Concentration while interacting with Reasoning Mind**

In cases where affective states may overlap you should use your best judgment and code the affect that is predominant. A good rule of thumb is to code the affect that corresponds with the student's primary behavior. If a student has gone *off-task* and they are *delighted* (or *frustrated* or *confused*) by their neighbor's behavior, code the affect related to their primary activity (in this case, their interaction with their neighbor). If, however, the student is doing enough work on their assignment to warrant an on-task behavior code, then the affective state of *engaged concentration* is probably more appropriate than their affective reaction to their friend. One possible exception to this rule of thumb occurs when you observe a student immediately after he or she has gone *off task*. If the student is still experiencing the residual effects of the educationally relevant affect that drove him or her *off task* (e.g. *boredom*, *confusion* or *frustration*), you may consider coding that emotion rather than the affective state related to the current social interactions he or she is having.

There will be times where it may seem like you have too little evidence to make a decision. If you're really not sure, you should code a "?" for that student's affect, but often you just need to pause and give yourself a few more seconds to take in the context of the affective state. Learning to do this can be challenging for novice coders. When you take extra time to code, it is important to remember that you are not looking for the next, easier affective state to categorize. Instead, you are trying to contextualize what you've seen. This can be challenging when students are cycling quickly through many emotions, so it is important to self-enforce the 20 second time

limit. If you cannot make a decision about the first affect you saw, code a question mark and move on to the next observation.

Coding for affective states can become particularly challenging when students are interacting with another person, particularly with another peer. Under these conditions, students are more likely to exhibit an affective state that does not fit a typical BROMP coding scheme, and in cases of overlap those affective states might be more prominent than anything in the BROMP coding scheme. (Although see Appendix B for other coding schemes.) Since the most prominent affective state observed is the one that should be recorded, you should consider using a '?' in these conditions, particularly if the student is off-task. If you find that an affective state is consistently reoccurring and it is not in your coding scheme, you should consider modifying your coding scheme before your next trip into the field. However, it is generally unwise to include categories that are unnecessary as they can skew inter-rater agreement estimates (Perreault & Leigh, 1989). (See Chapter 2 for more information on inter-rater agreement.)

It is also a good idea to spend a few minutes watching students as they come into the classroom. Some students may have a "resting face" (or baseline) that appears happier/sadder/more confused/etc. than others. These informal observations can help you decide how quickly you can safely make decisions about affect once the BROMP observation session begins.

Some new coders are particularly worried about optimizing their position relative to the student being coded, and this may be particularly important when coding affect. There is some research that suggests that the left side of a person's face contains more visual cues to emotional responses than the right (Mandal, Asthana, Madan & Pandey 1992; Skinner & Mullen 1991), although this effect may vary either by culture (Mandal, Harizuka, Bhushan & Mishra 2001; Rhodes & Lynskey 1990; Elfenbein, Mandal, Ambady, Harizuka & Kumar 2004; Mandal 1996) or by other significant social categories (Smith 1998). This effect may also be particularly susceptible to whether judges are being asked to look at static or dynamic images (Stringer & May 1981). At this time, we do not believe that there is adequate evidence to require BROMP observers to observe from the left, and trying to be consistent about such things may in fact result in reduced validity of BROMP coding. If the student is seated so that the left side of his or her face is against a wall, for example, it is better to code from the right than to risk getting so close to their physical space as to disrupt the classroom environment.

Furthermore, contextual clues, which are typically highly controlled for in laboratory settings (e.g., can be ignored) are often as important to appropriately coding affect as the student's own vocalizations and movements. (See Section 3.3.3.) For this reason, we recommend that observers focus their efforts on finding the space where they can get the best views of the student's posture, face, and computer screen *while remaining unobtrusive* rather than worrying about replicating optimal lab conditions. (Tips for remaining unobtrusive are provided below in Section 3.4.)

### 3.3.3 Examples of Ambiguous Affect:

1. A student who is experiencing "engaged concentration" MAY scowl. Particularly if facial expression is your primary clue for this student's affect, you should take a few seconds to observe the student before jumping straight to deciding they are confused or frustrated.
2. Confusion and frustration may occur at the same time, since one may trigger the other. If you see this, code the one that seems most prominent.
3. Yawning is a good indication of boredom, but it may just mean that the student is tired. Look for what they do before and after yawning.
4. Many people smile when they are frustrated (but it is not a smile of genuine happiness, such as the Duchenne smile).
5. Leaning back or changing postures can be indicative of certain behavioral and affective constructs, but they may simply indicate physical discomfort. Observers should be particularly aware of this in conditions where students are seated in furniture that may be larger or smaller than their frame.
6. Chewing gum, pencil tapping, and other repetitive behaviors can be revealing. Notice how much attention the student is paying to these tasks relative to whatever the teacher has assigned to them.
7. A student is laughing and telling their friend that the software is stupid, but seems to be doing so in order to hide feelings of being overwhelmed by the material. If you think they are frustrated by their inability to advance with the assignment, the laughing is probably irrelevant. Consider coding frustration or the more specific dejection, depending on your coding scheme.
8. A student who has been unable to complete a level in the educational game they were assigned goes off task, but still appears to be stewing with embarrassment, so you code dejection. When you return to the student for their next observation, they do not appear to have recovered, in that they are still off task, but they are now so engaged with their social activities that they have forgotten about their assignment. They are probably no longer experiencing dejection even though that is what has driven them to be off-task, so you should consider coding another affect instead.
9. You see a student go *off task* while you are coding one of their neighbors. When you arrive at their observation a moment later, they are giggling but twirling their mouse around by its cord and rolling their eyes in a way that indicates that they are still recovering from their boredom with the software. You should consider coding boredom. However, if the student is still off task when you return for the next observation, and they have found something interesting to entertain them, boredom is probably not the most appropriate code.
10. You see a student who is still answering questions within the software, but they are crying so hard that they are gasping for breath. While they may in fact be *on task* and focusing intently, profound *sadness* (not *engaged concentration*) is the predominant affective state. If you do not have a corresponding affect code for this category, a *?* is probably the best option.
11. If a teacher is answering questions that indicate a student does not understand part of the material, it might be evidence that the student is confused.
12. A student may be tapping rhythmically because he or she is engaged in concentration. Another student may be doing this because he or she is bored and off task. Use other contextual clues when interpreting such behavior.
13. At times, you may not be able to see a student's face because his or her hair is covering it. If the rest of the body language indicates that they are appropriately engaged in the material, you can code them as *engaged concentration*, but if you're in doubt, use the *'?'* code.

14. A student may be *off-task*, suggesting *boredom* with the system, but appear quite engaged in the doodles he or she is drawing. In general, we code this as *engaged concentration*, since BROMP coding schemes treat *off-task* instances of any construct separately from *on-task* instances of the same construct. If your research team has decided to treat this differently, please see the reporting requirements below.
15. A student is talking to his or her neighbor and appears to be confused, but the more you watch, the more you realize that he or she is flirting with the student in the next seat. If you believe the student you are coding actually understands the material and is simply posturing for their peer, confusion is probably not an appropriate code.
16. When in doubt, always use the "?" option.

## 3.4 Your Physical Presence in the Classroom:

1. **Talking to teachers.**
   - In general, you want to avoid talking to teachers during the observation process, but it is a good idea to talk to them before the students arrive. If you can, let them know how you will explain your role (if asked!) to the students before you get to the classroom.
   - You cannot (and should not) always prevent a full-introduction to the class if this is how teachers at that school regularly handle outside guests, but you should let teachers know that you prefer to keep this sort of interaction to a minimum so that you can remain unobtrusive. If you have this conversation with teachers ahead of time, they can provide students with consistent answers about your presence instead of introducing you as the person who is watching whether students are behaving!
   - One field coder used to tell teachers ahead of time, "If I'm doing it right, the students will think I'm staring angrily at my phone for the whole class." (Teachers are generally happy to help with this strategy and will sometimes volunteer critiques of how frustrating your phone seems to be.)
2. **Be unobtrusive by being nice, but boring.**
   - Kids are naturally curious, but they're also used to being watched. If you do not draw attention to yourself (and you are not staring at them), they will probably forget you are there.
   - Boring people attract very little attention. The best way to be boring is to look bored.
   - Smiling is not boring. It makes people wonder what you're up to.
   - As important as it is to be unobtrusive and boring, you do not want to look hostile or intimidating, even when you are entering or leaving the classroom.
   - A face that looks clueless but friendly (*not* interesting or happy) will make you look less threatening, particularly when you cannot avoid interacting with students (e.g. when you are collecting student log-in information or if one of them sticks their hand in your face to wave at you on the way by).
   - That said, looking like you are mildly annoyed at your phone is also OK, particularly during coding activities.
3. **Position and Movement.**
   - When picking places to stand or walk, remember that you want to use side glances and peripheral vision as much as possible.
   - Try to notice BOTH what is going on with the student's physical presentation and with the

computer screen. (If they are highly *engaged* with a videogame instead of the assigned educational software, they should be coded as *off task*.) However, if you can't see all three, facial expressions and body-language are generally more critical than the computer screen.

- In general, it is a good idea to stand diagonally behind the student you're coding. Typically we advise that you stand behind the student that you just observed while making the observation of the next student. However, you may be less obtrusive when you aren't moving at all.
- If you can find a location in the classroom that requires minimal movement but still allows you to see what's going on with large numbers of students, that is just as good as the technique where you're moving after each observation.

4. **Talking to Students.**
   - After you have collected names or login information, it is better to avoid talking or interacting with students as much as possible.
   - If a student asks you what you're doing, a truthful and non-problematic answer is, "We're looking at how students are responding to the software."
   - Students may sometimes come to you with complaints about the software. It is probably a good idea to validate their concerns with an apology even if you are not affiliated with the software developers (e.g., "Really? I'm sorry!"), but make it clear that you are not the developer of the software yourself. In these circumstances, your goal is to appear friendly and clueless (because you are). If it seems like a particularly important problem (e.g., inability to login), make sure that they've notified their teacher, but keep your interaction with the student as short as possible so that you can deflect attention away from yourself.
5. **Sensitization.**
   - If a student notices that they are being coded, it is best to abandon that observation (marking "?", "?" as discussed below).
   - Occasionally, one student will become sensitized to the observations and will repeatedly look at the observer to see if they are being watched. In this case, it is usually best to drop the student from observation for the rest of the session, marking the student as "?", "?" in all subsequent observations.
   - Smiling while you are coding should be kept to a minimum (it doesn't looking boring), but particularly if a student notices you, a quick smile can help you to appear less threatening and may prevent further sensitization. Be sure to disengage as quickly as possible, though. Staring out the window for a few seconds after such an encounter can help to re-establish your "bored, disinterested observer" status.
6. **Unusual Conditions.**
   - Fire drills, broken heating and cooling systems, and loudspeaker announcements cannot be avoided, but you should record them in your field notes.
   - You are not there to interact with the students or to report on them to the teacher. However, if you see a student doing someone particularly dangerous, use your best, adult judgment.
7. **Tech Support.**
   - During field observations of educational software, it can be useful to have a third person (non-observer) to entertain questions and inquiries, if possible. This is particularly useful when new software designs are being piloted.
   - The observer should never handle tech support. Not only is this distracting to the observer, who should be focusing solely on the observation task, but it also increases students' sensitization to the observer.

8. **Coding Teachers.**
   - In some cases, it may be important to code the teacher's behavior also. (Coding schemes for this are provided in Appendix B, Section B.4.)
   - Although HART allows you to use all three coding schemes at once, coding students is an intensive process. It is usually better to have more than one observer when considering teacher behaviors: one to code students and another to code teachers.
   - In some ways, it is more difficult to avoid observer effects for teacher codes, but if you are respectful and deferential, you can help to mitigate this problem.

## 3.5 Tips for Preventing & Correcting Miscoding:

- Seating charts aren't necessary, but they can facilitate the coding process.
- Having a seating chart as a reference allows a coder to cross-out students who leave early, come in late (and therefore weren't entered into HART at the beginning of the session), switch seats, complete the software task for the day, etc. It also allows the coder to make other notes as necessary.
- Obtaining a seating chart from a teacher ahead of time can speed the process of inputting student logins considerably, though it is important to confirm that no student is absent and that nobody has changed seats. It is also important to be cautious to verify that the teacher's seating chart accurately matches the classroom layout and that the students have not decided to switch chairs that day.
- Errors may occur during coding. If you realize that you have accidentally miscoded a student, you should note the observation number and the correct code for later correction of the data file. (If you have no paper to document this, we recommend noting the observation number and deleting the observations rather than trying make corrections.)
- If you are making many coding errors, slow down! It is important to enter data promptly, but not at the cost of entering incorrect data.

# Chapter 4: Human Affect Recording Tool (HART)

## 4.1 HART Overview and Documentation:

The *Human Affect Recording Tool*, or HART, is an application developed for the Android platform to facilitate BROMP observations (see Baker et al., 2012). HART was written using Java and synchronizes field observations to internet time, so that BROMP data can be precisely synchronized to the log file data from the educational software being studied. HART is available for direct download at http://www.columbia.edu/~rsb2162/bromp.html, and is also available through the US Army Research Labs GIFT (Generalized Intelligent Framework for Tutoring).

The HART application will ask you, the coder, to input data about the school and classroom you are observing. It will then ask you to enter student IDs for the students you are observing. Once you have done that, it will present those student IDs to you in the order that you entered them, asking you to code behavior and affect for each student until you tell it that you are finished. (See "HART Quick Start" section, below, for more details on this process.)

HART automatically saves your data to your Android device. Should the HART application unexpectedly crash or should you otherwise have non-catastrophic problems with your phone (e.g. a dead battery), your data will still be preserved. As an added measure of protection (and convenience), you can choose to email your data to a pre-designated address at the end of each class observation session.

Data can be harvested directly from the phone by connecting it to your computer and toggling the phone's settings to allow it to be accessed as a USB drive. Data is stored as a text file in a folder entitled "HARTdata," and you can safely copy your observation files to your computer without removing them from this location through the drag and drop function.

## 4.2 Adding Coding Schemes to HART:

New coding schemes can be added via editing the HartSchemas.xml file on your computer. You can edit this file using any text editor, including Notepad, Notepad++, or XEmacs. (Editing in Microsoft Word is NOT recommended). When editing the file, make sure to follow the exact same format as the file is currently in, include brackets and quotation marks, and save it with the same name. (You may want to save a backup first).

To add a new coding scheme, simply add a new "codelist", as shown in Figure 4, using the same format as shown there. You can add your new coding scheme either to behavior, affect, or intervention, and it will show up in that menu.

The first line of your new coding scheme gives the name of the coding scheme. For example, the first line of coding scheme "QED_BEHAV_V2" is written <codelist id="QED_BEHAV_V2">. The last line ends the coding scheme, and is written </codelist>. In between are the specific codes in the coding scheme. They appear in the HART app in the same order written in the file. For instance, "ACTIVE PARTICIPATION" will appear first in the list, "PASSIVE PARTICIPATION" will appear second in the list, and so on.

After you finish editing your HARTSchemas.xml file, connect your computer to your phone, and copy the new HARTSchemas.xml file to your phone's HART Data folder. Quit HART and restart it. If you don't know how to do this in Android, you can just restart your phone. Your new schemas should be ready to go!

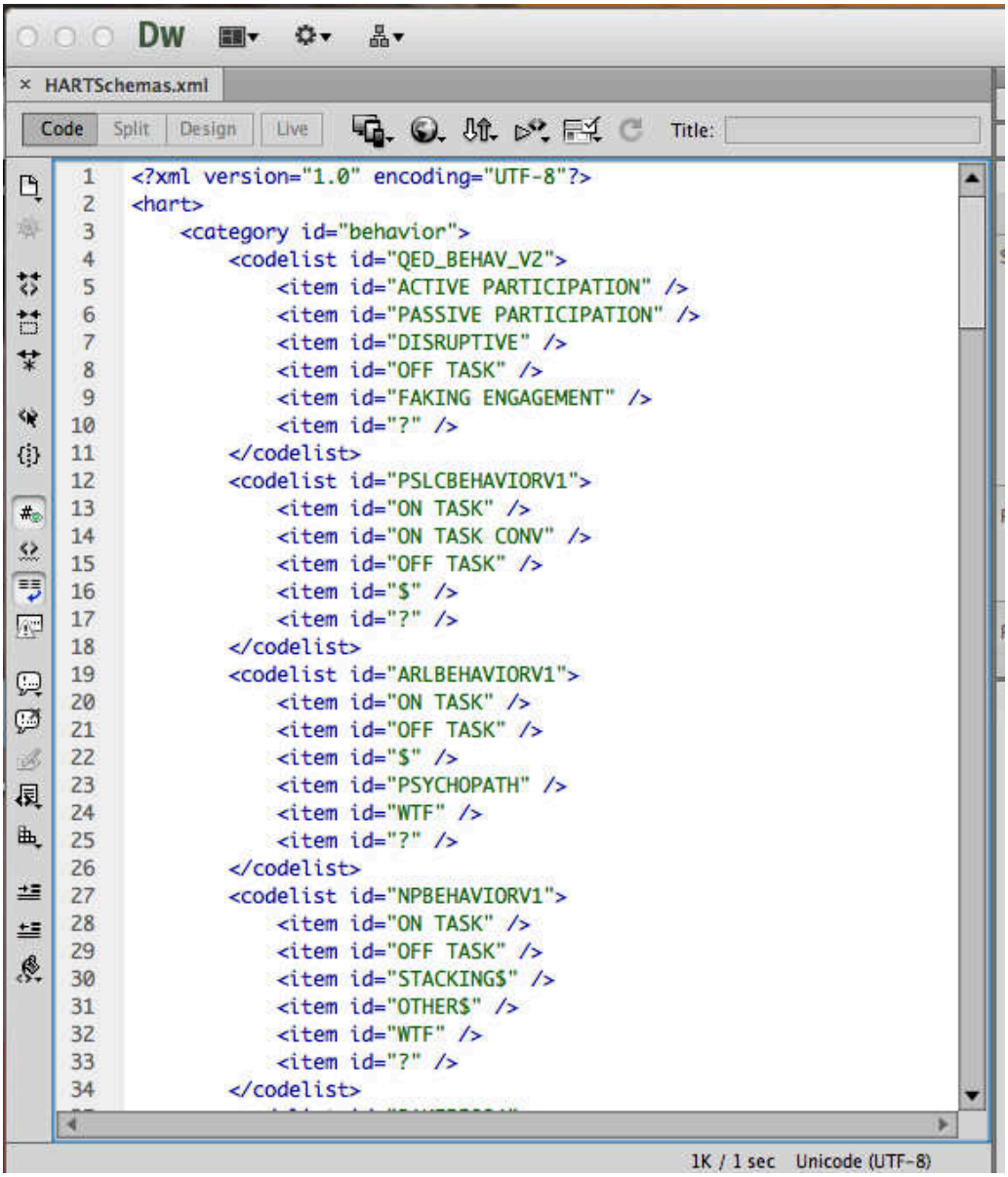



**Figure 4: Adding Coding Schemes to HART.**
This figure shows several of the current behavioral coding schemes available in HART. Additions and modifications to this list can be made easily made using a text editor.

## 4.3 HART Quick Start:

This section will provide you with a basic understanding of the mechanics of HART, the Android application developed to implement the BROMP 2.0 procedure for QFOs.

1. **Before opening HART, make sure that your device is not in airplane mode.**
   - If the device is in airplane mode, observations will not synchronize to internet time; this is not a problem if you are not synchronizing to software log files.
   - HART now gives you a warning message if you are in airplane mode, but it is easier to take care of this before you begin.
2. **Open HART** on your Android device. It will then provide prompt you through the following sequence.
3. **Select MODE:**
   - As of HART 8.8, you can use this option to choose to code using one, two, or three coding schemes (e.g. just behavior, just behavior and affect, or behavior, affect and an intervention).
4. **Input information about the coding session.** This window provides textboxes for the following information:
   - **USER:** that's you!
   - **SOFTWARE:** what software or learning context are students using?
   - **SCHOOL:** name of the school
   - **CLASSROOM:** typically [TEACHERNAME+CLASSPERIOD]
5. **Input the password:** "**maria**" (in all lowercase letters)
6. **Input the number of students** you will observer in this observation session.
   - You provide this number **before entering student IDs (or pseudonyms) and you, can NOT add or subtract** from that number once you tell the app to "start recording."
   - This can complicate things if students come in late or leave early. This can also be a problem if—in an effort to save time—you enter user-IDs before students arrive. (Invariably, someone will be absent when you do this.).
7. **Select the coding scheme(s)** you will be using.
   - Currently there are several pre-programmed into the application, but it is possible for a programmer to customize these to fit new needs.
   - Customization cannot be done in the field, nor is this typically advisable (although see Ocumpaugh et al., 2014).
   - **Appendix B** outlines current coding schemes to help with your selection process.
8. **Enter "student IDs" (or pseudonyms):** After you have provided the information about the fieldwork environment, you will be prompted through a series of windows, one for each student you will be observing.
   - When BROMP is being used to observe students using educational software, it is customary to enter the login ID that each student uses to access that software.
   - You will need to collect the login information from each student in the order that you intend to use for observations. Usually, this means that you must ask each student for their ID is at the beginning of the observation session.
   - If you do not need identifiable information about specific participants, you can enter nonsense strings or physical attributes (e.g., *green shirt*) that help you remember which participant you are observing.

9. **Synchronize and START RECORDING:**
   - Once you have entered in all of the students' login information, HART will ask you if you're ready to **start recording.**
   - You are, but you should check the "**synchronization**" box before you hit the "start recording" button, if you want to synchronize your observations to internet time. This procedure ensures that your device is still connected to the internet so that each observation is accurately synchronized to internet time.
   - If your device is NOT connected to the internet, you will receive a warning message. You can temporarily exit the problem to change this setting on your phone if you need to.
   - If your device IS connected to the internet, you will simply advance to first observation screen.
10. **Start Entering Observations:**
   - When you start recording, HART will present the "student ID" of the first student you entered. **Use the drop down menus on this screen to record your observations** in accordance with your training.
   - When you are satisfied with your coding selections, hit "**ok**" to move on to the next student. HART will automatically present students in the order that you entered them in at the beginning of the session.
11. **HART automatically generates information about each observation screen in order to improve accuracy**. This screen, which is shown in Figure 5, includes:
   - **Student Number:** corresponds to the order in which each student ID was entered into HART.
   - O**bservation Number:** calculates the number of observations *completed so far*.
   - C**ountdown clock:** helps you self-enforce any time constraints for each observations. BROMP guidelines limit the amount of time you should spend deciding how to code a given student (20 sec. limit), but classroom conditions can change, affecting coding patterns. Therefore, **nothing will happen** if for some reason the clock expires.
12. **Ending an Observation Session/Saving Files:**
   - When you are done, you should hit the **finish** button at the bottom of the observation screen.
   - HART will ask you to **confirm** that you are finished. Then it will advance you to a screen where you can send the data to a *pre-determined* email address.
   - If something malfunctions with the email process (or if the program crashes before this happens), your data will still be saved on the phone. You'll just have to retrieve it manually (via a USB connection, see Section 4.1).

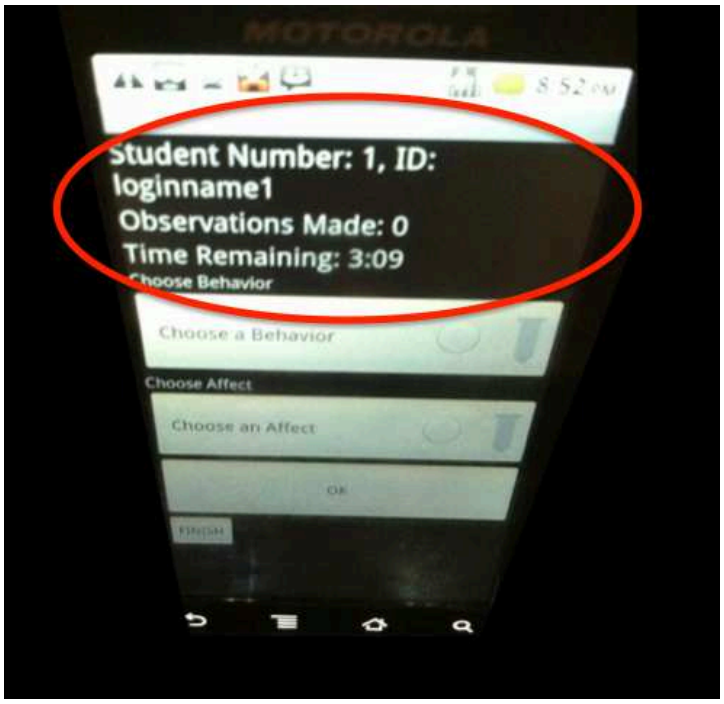
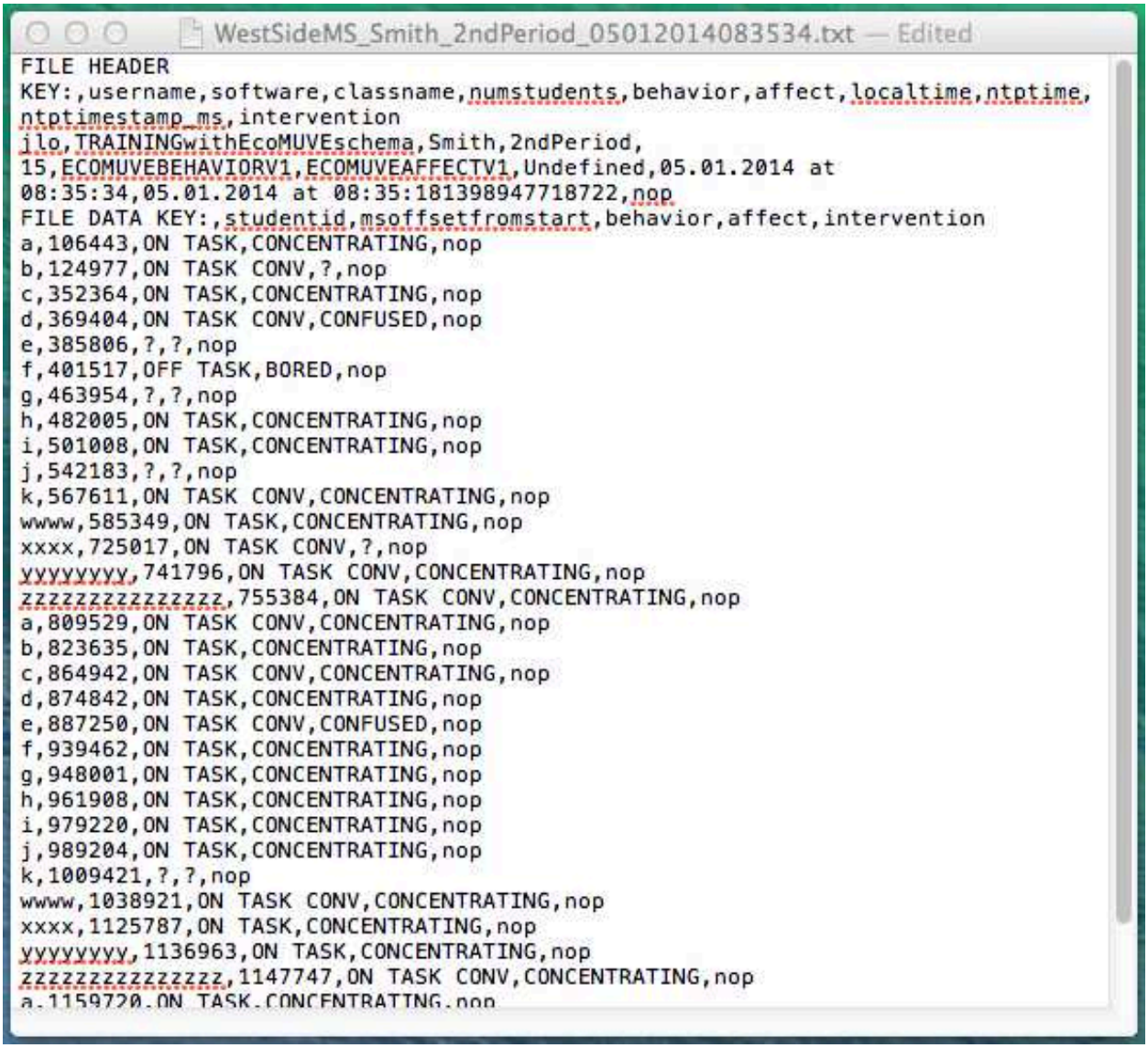



**Figure 5: A Student Observation Window in HART.**
Four pieces of information are given at the top of each information screen: (1) *Student Number*, which is automatically generated by HART, (2) *ID*, which displays what you entered about each student (usually login information, name, or pseudonym), (3) *Observations Made*, which is automatically updated each time you enter an observation, and (4) *Time Remaining*, a countdown clock that helps you to keep track of how much time you have taken per observation. Note that in this *hypothetical* example, we have entered "loginname1" as the studentID and the zero for *Observations Made* shows that we have not started recording data yet.

## 4.4 Using HART files:

HART saves observations as a text file, in comma-separated values format. An example of this is shown in Figure 3. HART.txt files can then be opened within Excel by importing it as a comma delimited file (e.g., Table 3). They can also be imported directly into other spreadsheet software and into most statistical software packages, using a similar process.



WestSideMS_Smith_2ndPeriod_05012014083534.txt — Edited

```
FILE HEADER
KEY:,username,software,classname,numstudents,behavior,affect,localtime,ntptime,
ntptimestamp_ms,intervention
jlo,TRAININGwithEcoMUVEschema,Smith,2ndPeriod,
15,ECOMUVEBEHAVIORV1,ECOMUVEAFFECTV1,Undefined,05.01.2014 at
08:35:34,05.01.2014 at 08:35:181398947718722,nop
FILE DATA KEY:,studentid,msoffsetfromstart,behavior,affect,intervention
a,106443,ON TASK,CONCENTRATING,nop
b,124977,ON TASK CONV,?,nop
c,352364,ON TASK,CONCENTRATING,nop
d,369404,ON TASK CONV,CONFUSED,nop
e,385806,?,?,nop
f,401517,OFF TASK,BORED,nop
g,463954,?,?,nop
h,482005,ON TASK,CONCENTRATING,nop
i,501008,ON TASK,CONCENTRATING,nop
j,542183,?,?,nop
k,567611,ON TASK CONV,CONCENTRATING,nop
wwww,585349,ON TASK,CONCENTRATING,nop
xxxx,725017,ON TASK CONV,?,nop
yyyyyyyy,741796,ON TASK CONV,CONCENTRATING,nop
zzzzzzzzzzzzzzz,755384,ON TASK CONV,CONCENTRATING,nop
a,809529,ON TASK CONV,CONCENTRATING,nop
b,823635,ON TASK,CONCENTRATING,nop
c,864942,ON TASK CONV,CONCENTRATING,nop
d,874842,ON TASK,CONCENTRATING,nop
e,887250,ON TASK CONV,CONFUSED,nop
f,939462,ON TASK,CONCENTRATING,nop
g,948001,ON TASK,CONCENTRATING,nop
h,961908,ON TASK,CONCENTRATING,nop
i,979220,ON TASK,CONCENTRATING,nop
j,989204,ON TASK,CONCENTRATING,nop
k,1009421,?,?,nop
wwww,1038921,ON TASK CONV,CONCENTRATING,nop
xxxx,1125787,ON TASK,CONCENTRATING,nop
yyyyyyyy,1136963,ON TASK,CONCENTRATING,nop
zzzzzzzzzzzzzzz,1147747,ON TASK CONV,CONCENTRATING,nop
a,1159720,ON TASK,CONCENTRATING,nop
```

**Figure 6: Example HART.txt file.**
This HART.txt file was generated during a BROMP training session. As it contains no identifying information, it is provided here as an example. Observational data from this figure are also given below in Table 3, to demonstrate what this information looks like in an Excel format.

File names are automatically generated by HART using two pieces of observer-inputted information (the name of the school and the name of the class) as well as automatically generated information. In Figure 6, you can see that the name of the school is *West Side MS,* and the class being observed was named *Smith 2ndPeriod*. The filename also supplies the date and start time of the observation session in month plus day plus four-digit year plus time format. (For example, the *0501201483534* found at the end of this file name means that the observation session was started on *May 01, 2014*, *34 seconds* after *8:35AM*.)

The data from the .txt file shown in Figure 6 was generated by a certified observer during a training session. The second row in the file contains information about the observation session. This includes information entered by the BROMP trainer that day as well as information that was automatically generated by HART. Labels for each piece of information appear above each piece of information in row 1, but in early versions of HART, the label *FILE HEADER KEY* causes the labels to be off-set by one from their corresponding information. Careful readers will see that Figure 6 contains the following **labels** and *information:* (1) **username**: *jlo*, (2) **software**: *TRAININGwithEcoMUVEschema*, (3) **numstudents**: *15*, (4) **behavior**: *ECOMUVEBHEAVIORV1,* (5) **affect**: *ECOMUVEAFFEECTV1*, (6) **localtime**: *undefined*, (7) **ntptime**: *05.01.2014 at 8:35:34*, (8) **Ntptimestamp_ms**: *05.01.2014 at 08:35:181398947718722*, (9) **intervention**: *nop*. Items 1-3 are entered by the observer (trainer) using the keypad on the phone. Items 4, 5, and 9 refer to drop down menu selections from various coding schemes; in this case, they used behavior and affect coding schemes that were developed for Dede's EcoMUVE software (Metcalf et al., 2011), but they did not use the third coding scheme, which is typically only employed in studies of interventions.

In order to make the information from the HART.txt file in Figure 6 more readable, the student observations from that file are also shown in Table 3, below. Here, we see that rather than entering identifiable information (e.g., the login information for the software students were using that day) the trainer entered pseudonyms for each of the 15 students observed (specifically: *a, b, c, d, e, f, g, h, i, j, k, wwww, xxxx, yyyyyyy,* and *zzzzzzzzzz).* These are given in the first column, followed by a timestamp in the second column. This timestamp is generated automatically and represents the number of milliseconds between the time at which the file was started and the time of each observation. The constructs coded for behavior, affect, and interventions follow in each row. In this case, observers did not use an *intervention* coding scheme, so *nop* (short for *no operator*) is given in each row.

**Table 3: Example Observation Data from a BROMP Training Session**.
Note that because this was a training session where names/login information were not collected, most students received single letter pseudonyms (e.g., *a* or *b*) that appear in the column labeled *studentid*. Other columns include a time stamp, a behavior code, and an affect code. Because an intervention (teacher behavior/classroom activities) coding scheme was not used that day, those observations are automatically filled with *nop*.

| studentid | msoffsetfromstart | Behavior | affect | intervention |
|---|---|---|---|---|
| a | 106443 | ON TASK | CONCENTRATING | nop |
| b | 124977 | ON TASK CONV | ? | nop |
| c | 352364 | ON TASK | CONCENTRATING | nop |
| d | 369404 | ON TASK CONV | CONFUSED | nop |
| e | 385806 | ? | ? | nop |
| f | 401517 | OFF TASK | BORED | nop |
| g | 463954 | ? | ? | nop |
| h | 482005 | ON TASK | CONCENTRATING | nop |
| i | 501008 | ON TASK | CONCENTRATING | nop |
| j | 542183 | ? | ? | nop |
| k | 567611 | ON TASK CONV | CONCENTRATING | nop |
| wwww | 585349 | ON TASK | CONCENTRATING | nop |
| xxxx | 725017 | ON TASK CONV | ? | nop |
| yyyyyyyy | 741796 | ON TASK CONV | CONCENTRATING | nop |
| zzzzzzzzzzzzzzz | 755384 | ON TASK CONV | CONCENTRATING | nop |
| a | 809529 | ON TASK CONV | CONCENTRATING | nop |
| b | 823635 | ON TASK | CONCENTRATING | nop |
| c | 864942 | ON TASK CONV | CONCENTRATING | nop |
| d | 874842 | ON TASK | CONCENTRATING | nop |
| e | 887250 | ON TASK CONV | CONFUSED | nop |
| f | 939462 | ON TASK | CONCENTRATING | nop |
| g | 948001 | ON TASK | CONCENTRATING | nop |
| h | 961908 | ON TASK | CONCENTRATING | nop |
| i | 979220 | ON TASK | CONCENTRATING | nop |
| j | 989204 | ON TASK | CONCENTRATING | nop |
| k | 1009421 | ? | ? | nop |
| wwww | 1038921 | ON TASK CONV | CONCENTRATING | nop |
| xxxx | 1125787 | ON TASK | CONCENTRATING | nop |
| yyyyyyyy | 1136963 | ON TASK | CONCENTRATING | nop |
| zzzzzzzzzzzzzzz | 1147747 | ON TASK CONV | CONCENTRATING | nop |
| a | 1159720 | ON TASK | CONCENTRATING | nop |
| b | 1180681 | ON TASK CONV | BORED | nop |
| c | 1197398 | ON TASK | CONCENTRATING | nop |

# Chapter 5: Reporting Standards

## 5.1 Importance of Reporting Standards

In order to maintain BROMP standards, it is important that both BROMP observers and BROMP developers maintain careful records. We understand that journal reviewers, page length restrictions and other publication requirements can be quite constraining, but researchers should make every effort to meet these reporting standards whenever possible. In this section, we outline *guidelines for publications* that report BROMP data, including the most important requirement:

**Only codes gathered from observers who have been certified in the BROMP method may be reported as BROMP observations. BROMP certification in one country does not qualify you to conduct BROMP observations in another country.**

We also urge you to help us to maintain our records about the use of BROMP by *reporting to developers*. In this way, we can ensure that BROMP standards are being met, increasing the value of your research.

## 5.2 Reporting to Developers:

**The authors of this training manual also respectfully request that certain information be emailed directly to them as a condition of being certified and using BROMP. This includes the following:**

1. *Certification of new coders.* Only someone who is BROMP certified can provide BROMP training. Anybody who trains another observer in BROMP should notify us once certification is complete so that we can maintain up-to-date records about BROMP certification, for reporting to our funders. (Please note that you **must** be a BROMP-certified coder in order to train a BROMP-certified coder, and that BROMP certification in one country does **not** qualify you to train coders in another country.) Please help us to maintain a list of certified BROMP coders in our records as soon as possible.
2. *Changes to the coding scheme.* Please let the BROMP developers know if you develop a new coding scheme or add a previously undocumented construct to an existing coding scheme, so that we can harmonize our coding schemes and notify other researchers who may find your modifications useful. Please include a description of any new constructs, differentiating those from any similar constructs that have been used before.
3. *Publications using BROMP.* Please notify us when you have had a report accepted for publication so that we can maintain records about BROMP's applications. This will also help us to circulate information about new publications to the BROMP research community, including citations to your work on the BROMP website and in new editions of the BROMP training manual.

We would also request that certified coders report back any helpful information about coding under unusual conditions, relevant decisions about ambiguous presentations of behavior or affect, or the extension of BROMP to new coding domains. This helps us to maintain the standards and reliability of the method.

## 5.3 Publishing Requirements:

Researchers should reference the BROMP coding schemes used, making every effort to include a description of each construct in any publications (or a citation to a definition elsewhere).

1. It is highly recommended that summary data be maintained for each fieldwork effort and that notes about unusual classroom conditions (e.g. heat waves, broken heaters, fire drills, etc.) be included in field notes that are maintained by the researcher.
2. If used in conjunction with another coding scheme (e.g. an observation schedule of classroom activities or teacher behaviors) or with log-file data from educational software, details about the synchronization process should be reported.
3. The total number of observations should be reported. Whenever possible, please include both an aggregate number and the number per class/school/learning environment (i.e., in cases where more than one learning environment was observed).
4. Whenever possible, field observation rates (e.g., average prevalence estimates) for individual constructs should be reported in publications and documented in field notes.
5. Decisions to discard or to post-hoc recode certain data (e.g. "?"s or instances of "engaged concentration" that co-occur with off-task behavior) should be documented accordingly.
6. Every effort should be made to report the duration of each observation session, observations per student, the average rate of observations per student (how long it typically takes you to cycle back to the same student) and the rate of observations per class. (These details can be easily calculated from HART .txt files.)
7. If ambiguous cases are particularly prevalent, decisions about those should be kept in field notes and reported upon to the maximum extent possible during publication.
8. If a new construct is added to the coding scheme before fieldwork begins, we encourage you to discuss this process with the developers ahead of time, whenever possible. Details about the cues typically associated with that construct should be documented and reported once fieldwork is completed.
9. If a new construct is added during fieldwork (e.g., Ocumpaugh et al., 2014), details about the process used to make that addition should also be documented and reported in research publications. (We do *not* recommend allowing recently certified coders to attempt this process. Changing coding schemes in the field is also *not* advisable for researchers using BROMP to study prevalence directly from the field recordings, as these changes may impact the estimates you are reporting in unpredictable ways.)
10. Because behavior and affect are culturally constrained and constructed, it is NOT advisable for a BROMP certified coder to observe students from a radically different cultural background than that of the group he or she was trained upon. Researchers who follow this advice should make efforts to mention this standard in their publication whenever possible (e.g., "Coders in this study were BROMP trained and certified on a culturally-similar population"). Researchers who would like to adapt BROMP to a new cultural condition should consult with the developers of this manual, who have already successfully accomplished this in three countries. We'd be very happy to partner in adapting BROMP for additional countries.

# Appendix A: BROMP Construct Descriptions

## A.1 Affective Categories (Commonly Used)

**Boredom:** the student appears to find the activity they are engaged with dull or tedious, often expressed as complete disengagement from the activity. Merely being off-task does not indicate boredom; in fact, off-task behavior often re-engages students (Moore et al., 2011; Sabourin et al., 2012). A student can be bored and fiddling with their computer mouse in a way that indicates they are starved for stimulation, but a student who is twirling the mouse around and giggling at the funny sounds his or her chair makes when it rocks back and forth is probably no longer bored even if boredom with the assignment drove them to that activity.

Readers who are interested in a more in depth description of boredom, its antecedents, and its consequences should consider the work of Eastwood, Frischen, Fenske and Smilek, (2012), who define boredom as:

> [T]he aversive state that occurs when we (a) are not able to successfully engage attention with internal (e.g., thoughts or feelings) or external (e.g., environmental stimuli) information required for participating in satisfying activity, (b) are focused on the fact that we are not able to engage attention and participate in satisfying activity, and (c) attribute the cause of our aversive state to the environment.

Eastwood et al.'s (2012) review of qualitative research on the nature of boredom highlights the associations of boredom with feelings of constrained agency, displeasure, anxiety, sadness, and anger, which may cause the student to have trouble performing tasks and may lead to misperceptions about how long a task is taking. Readers may also be interested in classificatory schemes for boredom (e.g., Nett, Goetz, & Hall, 2011; Van Tilburg & Igou, 2012).

**Confusion:** the student looks like they are having difficulty understanding the class materials or whatever they are most prominently engaged with. If the student is on task, the confusion must appear to be related to their task. (If the student appears mildly confused about their neighbor's behavior, but is continuing to successfully remain on task, this does not qualify as task-related confusion, and engaged concentration is probably the most appropriate code. However, when in doubt, always use the "?" code.) Particularly among younger children in the U.S., this emotion is often quite easy to recognize, but researchers in the Philippines report more difficulty in coding for this category, possibly because it is less culturally acceptable to express confusion in cultures with greater power distance. Please note that cues for confusion do not have to be expressed facially. Coders can use verbal cues (e.g. asking someone for help or an explanation) or other conventional signals of confusion (e.g. a student raising his or her hand to try to get help).

**Delight:** the student is smiling or otherwise indicating that they are having a pleasurable experience. Remember, however, that not all smiling warrants a delight code. If the student giggles at something funny their friend says about the teacher's wardrobe, but is maintaining on-task concentration, then engaged concentration is probably the more appropriate code. Furthermore, smiling has also been found to correlate with frustration in several *in situ* studies (Hoque & Picard, 2011).

**Engaged Concentration:** the student is paying focused attention to their primary current task task (be that on-task assignments or off-task behaviors). Some students can multi-task while continuing to be in a state of engaged concentration. This is the affective state associated with Csikszentmihalyi's (1990) construct of *flow*. In early work, it was called flow (e.g. D'Mello et al., 2007; Rodrigo et al., 2007), but it was later renamed to avoid adopting other elements of Csikszentmihalyi's model. Some researchers simply refer to it as "engaged". Often students who are working individually scowl when presenting engaged concentration. If you are unsure whether the student is concentrating or frustrated, it is a good idea to take a little extra time to make your decision. Students who are frustrated will sometimes appear more dynamic than students who are concentrating.

**Frustration:** Frustration is coded when the student presents feelings of distress or annoyance, although some students may manage or interpret this annoyance differently than others. For example Gee (2007) has discussed *pleasurable frustration*, an affect that seems to present when students enjoy being challenged. Note also that *in situ* research has demonstrated that smiling is often a sign of frustration in the US (Hoque & Picard, 2011).

**Surprise:** coded when posture, facial expressions, or vocal expressions indicate that a student's previous affective state was interrupted unexpectedly.

## A.2 Affective Categories (Less Commonly Used)

**Anger:** One of Ekman's basic emotions, anger differs from frustration in that it is a response to feeling threatened, either deliberately or otherwise. Its presentation is usually more intense than frustration, and it is more likely to be associated with belligerent behaviors. Students who are angry may change their volume and physical presentation from one extreme or another (very quiet vs. very loud, or very withdrawn vs. very aggressive), and their behavior may be marked by a lack of civility, particularly in students with poor emotional regulation.

**Confrustion:** this affective category was first coined in Liu, Pataranutaporn, Ocumpaugh, & Baker, (2013), which showed that patterns of confusion and frustration better predicted learning outcomes when treated as the same construct. Combined with field observations, we have reason to believe that particularly within some populations of students in the United States, confusion and frustration may feed on one another to the point that it is either not possible or not fruitful to distinguish between the two. This may be more likely among young students, who have less metacognitive awareness and are therefore less likely to be able to regulate unpleasant feelings like confusion. Please also note that this is culturally sensitive construct. BROMP coders doing fieldwork in the Philippines do not typically observe this relationship, and our partners in India have developed a coding scheme that uses disgust rather than frustration because the latter in considered socially inappropriate there. To our knowledge, it has not been observed in lab studies within the United States, either.

**Contempt:** though not typically recognized as an educationally relevant affective state in the United States, contempt was added to the BROMP coding scheme in India to accommodate the fact that demonstrations of contempt replace demonstrations of frustration in this culture, where frustration is not socially acceptable. Contempt is sometimes (though rarely) seen populations in the USA, although it is usually directed towards the learning environment.

**Dejection:** This is state of being saddened by failure. Like frustration, the student presents evidence of distress or feeling overwhelmed. Students experiencing dejection often seem embarrassed. In many cases, this emotion may co-occur with frustration and anger, but students often try to hide this emotion from their peers. They may exhibit significant social posturing in an effort to prevent other students from realizing that they are struggling with the material. Note that off-task behavior may be a strategy that dejected students use to self-regulate in order to overcome dejection.

**Disgust:** One of Ekman's basic emotions, disgust indicates that the student finds the real (or virtual) task "icky;" it was first used as a BROMP category during observations of EcoMUVE, an educational software that asked students to determine why the virtual fish in that environment were dead (e.g. Ocumpaugh et. al., 2014)

**Eureka:** A moment of sudden understanding or awareness, usually accompanied by indications of surprise. This category is common among early research on educationally relevant affective states (e.g., D'Mello et al., 2007), but it is quite infrequent in actual learning experiences (D'Mello et al., 2007; Lehman et al., 2008).

**Happiness:** One of Ekman's basic emotions, happiness is demonstrating contentment or other expressions of well being. (Although less intense than delight). Happiness is common in learning (Lehman et al., 2008) but has not yet been shown to lead to better learning outcomes. (There is research showing that affect with positive valance is associated with learning, but that work did not distinguish happiness from engaged concentration or delight).

**Pride:** demonstrating pleasure or satisfaction in accomplishment

**Sorrow/sadness:** One of Ekman's basic emotions. In sadness, sometimes referred to as sorrow, students appear unhappy or regretful. This emotion is also not typically coded in most BROMP coding schemes, but it can be educationally relevant when students are interacting in virtual worlds, for instance, when a virtual animal or learning companion dies.

### A.3 Behavioral Categories

**On-Task Behavior:** refers to a student who is doing what he or she is supposed to be doing. Typical coding schemes do not differentiate the different kinds of on-task behavior other than identifying when students are participating in conversation (see below) or working in isolation. However, it is also possible to make other distinctions:

> **Creative Metanarrative:** Occurs when a student is having a discussion about a learning environment, but seems to be inventing his or her own storyline. For example, a student who is telling his or her friends about the interactions between non-existent criminal elements and police officer in a virtual world that teaches about environmental science (Ocumpaugh et al., 2014).
>
> **On-Task Conversation:** refers to a student who is working towards his or her assignment while having a conversation with the teacher or another student about the subject matter or learning task.

**On-Task Giving/Receiving Answers:** A student who is focused on the learning task, but solely on what the answer is.

**On-Task Help Seeking:** refers to a student who has paused work, but only because he or she is seeking help from another student or the teacher

**On Task Procedural/Supplies:** refers to a student who has paused work in order to get supplies or address equipment breakdowns.

**Proactive Remediation:** a teacher receives information on student progress, and intervenes to work with the student (see Miller et al., in press)

**Off-Task:** refers to a student who is not working on the educational task assigned by their teacher. In many BROMP coding schemes, we do not distinguish among the many types of off-task behavior, but below are some codes that have been used.

**Aggression:** refers to a student who is not only off task but behaving in a threatening manor towards another student.

**Off-Task Passive:** refers to a student who is off task but not interacting with anybody or doing much of anything in particular. For instance, the student may be sleeping or staring into space.

**Off-Task Social:** refers to a student who is off-task but interacting with a peer

**Off-Task Supplies:** refers to a student who is not interacting with any of his or her peers, but who is playing with an object like a pencil or a computer mouse.

**Non-Passive Withdrawal:** refers to an off-task student who is not bothering anybody or playing with any objects, but who is not really being passive, either. This is a sort of conspicuous, attention seeking behavior that might include a student trying to defy a teacher by putting their head down on the desk and refusing to work.

**Gaming the System:** is a special behavior that is neither on-task nor off-task. It occurs when a student is still engaged with the learning software, but is not engaged with learning. Instead, they are attempting to advance through system without actually learning the material (Baker et al., 2004).

**WTF:** Refers to "Without Thinking Fastidiously" (Wixon et al., 2012). Any other meanings of the acronym are sheer coincidence. That's our story, and we're sticking to it. WTF behavior consists of actions within the learning environment that are not targeted towards the learning task or successful performance. Examples include placing virtual cactuses on top of virtual patients (Rowe et al., 2011), drawing smiley faces instead of plotting points, climbing buildings (Rowe et al., 2011), or repeatedly pausing and unpausing a simulation at high speed (Wixon et al., 2012).

**Psychopath:** Used in military simulations to indicate that students were initiating so-called "friendly fire" and shooting their teammates, or engaging other forms of aggressive behavior that were enabled (but not encouraged) by the software.

**Stacking:** Used to record a type of behavior specific to Newton's Playground, where students create objects that can be "stacked" on top of one another instead of designing machines as the software designers intended them to do. Stacking might be considered a kind of *gaming the system* that is specific to Newton's Playground.

# Appendix B: Current Coding Schemes

## B.1 Overview

Originally, HART was far less customizable. It required coders to use two coding schemes, one for behavior and one for affect. Now coders can choose to use one, two, or three coding schemes, selecting from schemes developed for three categories: behavior (Section B.2), affect (Section B.3, and teacher interventions (Section B.4). Currently developed schemes are provided in the tables in this section, but concrete descriptions are found in Appendix A.

## B.2 Coding Schemes for Behavioral Indicators of Engagement

The most commonly used coding schemes are those developed for the Pittsburgh Science of Learning Center (PSLC), shown both in Table 4 (below) and in Table 2 (Chapter 3). This was refined from the original BROMP coding scheme used in the original studies of *gaming the system*, which is denoted with a dollar sign in HART (Baker et al., 2004) and defined in more detail in Appendix A.

**Table 4: Primary coding schemes for behavioral constructs**

Coding schemes from Baker et al.'s (2004 Pittsburgh Science of Learning Center (PSLC) at Carnegie Mellon University. Note that $ is used to code gaming the system, and ? as an abbreviation for "other."

| Baker2004 | PSLC |
|---|---|
| On task | On task |
| On-task conversation | On-task conversation |
| Off-task solitary | Off task |
| Inactive | $ |
| $ | ? |
| ? | |

Since then, several other behavioral coding schemes have been developed, including those for the Army Research Labs' (ARL) work with vMedic, and Physic's Playground, formerly known as Newton's Playground (NP), EcoMUVE, and Refraction. Note that some of these studies necessitated behavioral constructs that were not used in previous studies. For example, *stacking* refers to a specific kind of behavior similar to gaming the system; it is unique to the Physics Playground environment. Readers should also note the UserDef categories in the EcoMUVE coding scheme. These were created because we were unable to run a pilot study with EcoMUVE, a virtual environment that we thought might induce interesting new behavioral constructs that had not been seen in previous research. In fact, we did use this system to code for a construct called *creative metanarrative*, which is defined in more detail in Appendix A.

**Table 5: Behavioral coding schemes developed for specific software**

| ARL | NP | EcoMUVE | Refraction |
|---|---|---|---|
| On task | On task | On task | On task |
| Off task | Off task | On-task conversation | On-task conversation |
| Psychopath | Stacking$ | Off task | Off task |
| WTF | Other$ | $ | Receiving Help |
| $ | WTF | UDef1 | Giving Help |
| ? | ? | UDef2 | ? |
| | | UDef3 | |
| | | ? | |

In addition to codes developed for specific software environments, codes have also been developed for observations in regular classroom. For example, Fisher's team at Carnegie Mellon University has been using BROMP to study engagement in kindergarten classrooms (e.g., Godwin et al., 2014). Readers should note that the some of the behavioral codes for the Fisher schemes are not listed in Appendix A. These codes, which can be seen in Table 6, all refer to the various types of off-task behaviors that children may engage in. (On task behaviors are coded with on-task.)

The largest user of BROMP, however, is now the QED project in India. Since the summer of 2014, they have been using BROMP to help teachers develop more engaging pedagogical strategies in an effort to improve educational outcomes for their students. To date, the QED coding scheme has not been used in the United States or the Philippines.

**Table 6: Behavioral coding schemes used in non-technological classrooms.**

| FisherD12011v4 | FisherD22011 | QED |
|---|---|---|
| On Task | Self | Active Participation |
| Peer | Environment | Passive Participation |
| Environment | Peer | Disruptive |
| Supplies | Walking | Off task |
| Self | On task | Faking Engagement |
| Walking | ? | ? |
| Other Off-task | | |
| ? | | |

## B.3 Coding Schemes for Affective Indicators of Engagement

Development of affective coding schemes, as explained in Chapter 1, was based on research about educationally relevant affective states. These constructs were not coded in the first research to use BROMP (Baker et al.'s 2004 study of *gaming the system*), but were added for later research being conducted by the Pittsburgh Science of Learning Center (PSLC). In general, we recommend that either this coding scheme or the one based on D'Mello et al.'s (2007) research. These schemes are shown in Table C.

**Table 7: Primary coding schemes for Affective States**

Coding schemes developed for the Pittsburgh Science of Learning Center and from D'Mello et al's (2007) work on educationally relevant affective states. Note that as in the behavioral coding schemes, ? is used as our "other" category.

| PSLC | D'Mello2007 |
|---|---|
| Boredom | Boredom |
| Confusion | Confusion |
| Concentration | Concentration |
| Frustration | Delight |
| ? = other | Frustration |
| | Neutral |
| | Surprise |
| | ? |

Despite the utility of the coding schemes in Table 7, it is sometimes useful to develop coding schemes that reflect the needs of a specific learning environment. This is perhaps even more important for affective coding schemes than for behavioral coding schemes. Tables 8 and 9 show the coding schemes developed for specific software platforms and for non-technological environments, respectively. Readers should note that as with the behavioral coding scheme for EcoMUVE (Table 5), the affective coding scheme contained a UserDef (User Defined) category that allowed us to add constructs in the field. This is typically not advisable for novice coders, particularly for studies where BROMP is not being used to develop software models of behavior and affect, but it did allow us to capture the construct *disgust,* an engaged reaction that many students had to the dead fish in that virtual environment. Readers may also be interested to learn that after a pilot study of Physics Playground (Newton's Playground), it was determined that many of the affective states listed here were too rare to include in field work. As a result, we relied upon the EcoMUVE coding scheme. This scheme already contained the construct *delight,* and it allowed us to use the UserDef option to code for *dejection.*

**Table 8: Affective coding schemes for specific software platforms**

Coding schemes that have been developed for the Army Research Labs' (ARL) work with vMedic, for Physics Playground, formerly Newton's Playground (NP), and for Refraction.

| ARL | NP | EcoMUVE | Refraction |
|---|---|---|---|
| Anxious | Anger | Boredom | Boredom |
| Boredom | Anxious | Confusion | Concentration |
| Confusion | Bored | Concentration | Confusion |
| Concentration | Concentration | Frustrated | Delight |
| Frustration | Confused | Delight | Eureka |
| Surprise | Curious | Sorrow | Frustrated |
| ? | Delight | UserDef | Surprise |
| | Excited | ? | ? |
| | Frustrated | | |
| | Happy | | |
| | Hope | | |
| | Pride | | |
| | Sad | | |
| | Surprised | | |
| | ? | | |

**Table 9: Affective coding schemes used in non-technological classrooms**

| Fisher | QED |
|---|---|
| Boredom | Boredom |
| Confusion | Confusion |
| Concentration | Focused |
| Delight | Enthusiastic |
| Silly | Mildly Interested |
| Other | Eureka |
| ? | Delight |
| | Contempt |
| | Disinterested |
| | ? |

## B.4 Interventions and Other Coding Schemes

Particularly when the behavioral and affective constructs are not being compared to students' interactions with an online learning environment, it is useful to document what kinds of activities a student is participating in when they are being coded or what kinds of interventions a teacher attempts. Although HART now permits the use of a third coding schemes, we generally recommend that a second coder be responsible for the coding of classroom conditions or teacher activities. In our experience, coding students is an intensive process and dividing attention across more than one person at a time is challenging. To date, most of the research involving a third coding schemes has been pioneered by Fisher's team at Carnegie Mellon University, who have typically used a second coder. We encourage readers to contact Fisher (or the developers of this manual) with questions about these schemes, which are shown in Table 10.

**Table 10: Interventions and other coding schemes**

| FisherD12011v4 | Fisherv1 | FisherClassActs |
|---|---|---|
| Intervention | Intervention | Wholedesks |
| None | Whole | Wholecarpet |
| | ? | Sgingdiv |
| | | Sgteach |
| | | Individual |
| | | Other |
| | | ? |

# Appendix C: Video Resources

We have a limited number of videos that were taken during BROMP training sessions that we sometimes use in Phase 1 of our training. Privacy restrictions prohibit us from making these publically available, but some people find it useful to spend time watching students before starting BROMP training. If you are unable to gain access to classrooms before you begin your training, it may be useful to familiarize yourself with the sorts of behaviors, postures, and facial expressions that are typical among students by looking at publically available videos of classrooms.

Finding videos for specific age groups, content, or conditions can be challenging. For your reference, you will find a list of YouTube videos that were available at the time this manual was published, with some information about the grade level of these students (Table 11). Please note that we are not affiliated with the makers of any of these videos and cannot make any comments on the veracity of any information presented either in the videos or in the comment sections that accompany them.

**Table 11: YouTube videos of naturalistic classroom conditions**

| Grade | Title | Web Address |
|---|---|---|
| PreK | Observing Young Children | https://www.youtube.com/watch?v=t1Xtr3RKjGc |
| PreK | A Westwood Preschool Classroom 2 | https://www.youtube.com/watch?v=0Egr2Xxr95k |
| PreK | Early Education Enrichment one hour class observation | https://www.youtube.com/watch?v=GSQFJipws4g |
| K | Teaching 21st Century Skills in Kindergarten | https://www.youtube.com/watch?v=NAfFnna9_MU |
| K | Assessment of Teaching and Learning: Classroom Observation | https://www.youtube.com/watch?v=_qyuSz0Y9GU |
| K-2 | Special Education K-2 Teacher | https://www.youtube.com/watch?v=an3ngVFbJC0 |
| 1 | Lost Lake Elementary Promo | https://www.youtube.com/watch?v=Gkeu3nQwLTE |
| 1 | Syracuse Academy of Science Charter School | https://www.youtube.com/watch?v=auCc_BTmuFc |
| 1 | "Whole Brain Teaching" | https://www.youtube.com/watch?v=aaweXw03kQI |
| 2 | classroom observation part 1 | https://www.youtube.com/watch?v=tAz7TD02ytU |
| 2 | Second Grade Computer Lab | https://www.youtube.com/watch?v=liVyB1_p5k4 |
| 2 | iPads in the Classroom | https://www.youtube.com/watch?v=IzSNdxsfk0Q |
| 2 | 2nd Grade Everyday Math Lesson 10.8 | https://www.youtube.com/watch?v=zgW9hJE_n_s |
| 3 | Elementary Math Classroom Observation | https://www.youtube.com/watch?v=jzq-kuyhiqs |
| 4 | Fourth Grade Guided Reading - Hawthorne Elementary - Mrs. Sorenson's Class | https://www.youtube.com/watch?v=Yw0fT3Lm0sY |
| E | Local Elementary Students Use Ipads In Classrooms | https://www.youtube.com/watch?v=8x_61o8dKYQ |

| | | |
|---|---|---|
| E | each Like a Champion: Getting everyone's attention in class | https://www.youtube.com/watch?v=EC0ltKOwF_A |
| E | Using the iPad to teach the math concept "before" | https://www.youtube.com/watch?v=vpvpR51v92E |
| E | Day School's Computer Lab | https://www.youtube.com/watch?v=IwWM7eEmaoc |
| E | Gloria Dei Lutheran School - Hampton Virginia | https://www.youtube.com/watch?v=phyiIPrGO-U |
| E | K-5 Computer Lab | https://www.youtube.com/watch?v=Zqzkyn1BfvY |
| E | Video of Lesson - Student Teaching | https://www.youtube.com/watch?v=P_atMXYwX3A |
| E | Clincy. Classroom Observation 4.18.12 | https://www.youtube.com/watch?v=xah_C_aTkZs |
| H | Classroom Observation | https://www.youtube.com/watch?v=yEj8yTEbSdE |
| H | High School History Lesson 2 | https://www.youtube.com/watch?v=q1EsaCzWIZ8 |
| M | Classroom Design | https://www.youtube.com/watch?v=UZh76TcDnSw |
| M-H | Classroom management - Week 1, Day 1 | https://www.youtube.com/watch?v=pgk-719mTxM |
| 6 | Mr. Mehney's 6th Grade Math Class.m4v | https://www.youtube.com/watch?v=xtzhdIR_Y7E |
| 6 | 6th Grade Resource class | https://www.youtube.com/watch?v=g-Cc0BWMrJY |
| 8 | Mr. Meriweather's 8th Grade Science Class | https://www.youtube.com/watch?v=jemtGiy2v2Q |
| 9 | Maggie Goldstein Classroom Observation (5.6.14) | https://www.youtube.com/watch?v=cATQhVhBXsc |
| 10 | Classroom Clips - 10th Grade Science - Steve Cornell (Part 1) | https://www.youtube.com/watch?v=tOWYMCmx_0c |

# Appendix D: Other Coding Schemes

There is a substantial body of research on classroom behaviors and students emotional states, but BROMP is somewhat unique in that it uses direct observation to record both simultaneously. When classroom observation systems became popular in the 1960s, most coded primarily based on behaviors, although a few (particularly Perkins, 1965) recognized the importance of classifying attentional data. Very few researchers within education have used direct observation to study emotion (although see Izard et al., 2007), instead preferring to use survey measures, many of which rely on self-report (see Table 12).

**Table 12: Survey Instruments from Previous Research**

| Acronym | Instrument | Citation |
|---|---|---|
| BPS | Boredom Proneness Scale | Farmer & Sundberg, (1986) |
| CCSSE | Community College Survey of Student Engagement | McClenney et al., (2012) |
| MDBS | MutliDimensional Boredom Scale | Fahlman et al., (2011) |
| MES | Motivation Engagement Scale | Reschly et al., (2014) |
| MMS | Me and My School | Darr et al., (2012) |
| NSSE | National Survey of Student Engagement | IUCPR, (2003) |
| SEFECS | Student Engagement and Family Educational Culture Survey | Leithwood & Jantzi, (2000) |
| SEI | Student Engagement Instrument | Appleton et al., (2006) |
| SESQ | Student Engagement in Schools Questionaire | Hart et al., (2011) |
| TERF-N | Teacher Engagement Report Form-New | Hart et al., (2011) |
| ZBS | Boredom Susceptibility Scale | Zuckerman et al., (1978) |

An exhaustive review of the observation systems used to study engaged behaviors in the classroom is beyond the scope of the current work, but readers might be interested in exploring some of the instruments that have been used by previous researchers, particularly if modifications to BROMP would help them to improve their own observational studies. A list of some of the currently published observational instruments for studying behavioral engagement is given in Table 13, and readers may also be interested in consulting review articles (e.g., Adamson & Wachsmuth, 2014; Anderson, 1981; Fredericks et al., 2011; Hintz et al., 2002; Hops et al., 1995; Nock & Kurtz, 2005; Riley-Tillman et al., 2005; Skinner et al., 2000).

**Table 13: Observational Instruments from Previous Research**

| Acronym | Name | Citation |
|---|---|---|
| AAE | Assessment of Academic Environments | Overton, (2004) |
| AET-SSBD | Academic Engaged Time of the SBBD | Walker & Severson, (1990) |
| APECP-RV | Assessment Profile for Early Childhood Programs: Research Version | Abbot-Shim et al., (2000) |
| APEEC | Assessment of Practices in Early Elementary | Hemmeter, (2001) |

| | | |
|---|---|---|
| | Classrooms | |
| AROS | Attending Round Observation System | Weinholtz et al., (1986) |
| ASEBA | Achenbach System of Empirically Based Assessment | McConaughy & Achenbach, (2009) |
| ASOS | Activity Setting Observation System | Rivera & Tharp, (2004) |
| BAG | Behavioral Assessment Grid | Cone, (1978) |
| BASC-2 BESS | Behavior Assessment System for Children-2 Behavior and Emotional Screening System | Kamphuas & Reynolds, (2007) |
| BASC-SOS | Behavior Assessment System for Children--Student Observation System | Lett & Kramphaus, (1992) |
| BASC-TRSC | Behavior Assessment System for Children—Teacher Rating Scale for Children | Baker et al., (2008) |
| BEEOS | Behavior and Emotion Expression Observation System | Izard et al., (2007) |
| BOSS | Behavioral Observation of Students in Schools | Shapiro, (2003) |
| C-COS | Child-Caregiver Observation System | Boller & Sprachman, (1998) |
| CAR | Classroom Activity Record | Evertson & Burry, (1988) |
| CASS | Classroom Assessment Scoring System | La Paro et al., (2004) |
| CASS | Classroom Assessment Scoring System | Pianta et al., (2008) |
| CBCL | Child Behavior Checklist | Achenbach, (1983) |
| CISSAR | Code for Instructional Structure & Student Academic Response | Greenwood et al., (1978) |
| CISSAR | Code for Instructional Structure & Student Academic Response | Stanley & Greenwood, (1981) |
| CLASS-S | Classroom Assessment Scoring System–Secondary | Casabianca et al., (2014) |
| Classroom AIMS | Classroom Atmosphere, Instruction/content, Management, & Student-engagement | Roehrig & Christesen, (2010) |
| Classroom CIRCLE | Classroom Code for Interactive Recording of Children's Learning Environment | Atwater et al., (2009) |
| COC | Classroom Observation Code | Abikoff & Gittelman, (1985) |
| COEMET | Classroom Observation of Early Mathematics Environment and Teaching | Sarama & Clements, (2007) |
| COI | Classroom Observation Instrument | Spears, (2013) |
| COP | classroom observation protocol | Harniss et al., (2007) |
| COPS | Classroom Oral Participation Scheme | King (2013) |
| COS | Classroom Observation System | Dotterer & Lowe, (2011) |
| COSMIC | Classroom Observation to Measure Intentional Communication | Brittain, (2012) |
| CPI | Classroom Practices Inventory | Hyson et al., (1990) |
| CPP | Classroom Performance Profile | Crosby & French, (2002) |
| CTRS-R | Conners' Teacher Rating Scale-Revised | Conners, (1997) |

| | | |
|---|---|---|
| DBR-SIS | Direct Behavior Rating-Single Item Scales | Chafouleas et al, (2010) |
| DOF | Direct Observation Form | Achenbach, (1986) |
| DOF | The Direct Observation Form | Volpe et al., (2009) |
| DOS | Dyadic Observation System | Good & Brophy, (1994) |
| EAS | The Emergent Academic Snapshot | Ritchie et al., 2001 |
| EBASS | Ecobehavioral Assessment Systems Software | Greenwood et al., (1994) |
| ECCOM | Early Childhood Classroom Observation Measure | Stipek & Byler, (2004) |
| ECERS | Early Childhood Environment Rating Scale | Harms et al., (1998) |
| ECERSE | Early Childhood Environment Rating Scale Extension | Sylva et al., (2003 |
| ELLCOT | Early Language & Literacy Classroom Observation Tool | Smith et al., (2008) |
| ESTEEM | Expert Science Teaching Educational Evaluation Model | Burry-Stock & Oxford, (1994) |
| FFT | Framework for Teaching | Danielson, (2011) |
| GOM | General Outcomes Measure | Christ et al., (2011) |
| HSOS | Hit-Steer Observation System | Reeve & Tseng, (2011) |
| ICP | Inclusive Classroom Profile | Soukakou, (2012) |
| inCLASS | Individualized Classroom Assessment Scoring System | Downer et al., (2010) |
| IPI | Instructional Practices Inventory | Hyson et al., (1990) |
| IQA | Instructional Quality Assessment | Junker et al., (2004) |
| ISTOF | International Systematic Teacher Observation Framework | Mujis et al., (2014 |
| ITM | Interactive Teaching Map | Kerr et al., (1985) |
| LAMM | Learner Activity Monitoring Matrix | Williams & Carvalho, (2010) |
| M-CBM | Mathematics Curriculum-Based Measurement | Christ & Vining, (2006) |
| MACOS | Mathematics Classroom Observation Schedule | Ndirangu et al., (2011) |
| MOOSES | Multi-option Observation System for Experimental Studies | Tapp et al., (1995) |
| MQI | Mathematical Quality of Instruction | University of MI, (2006) |
| MS-CISSAR | Mainstream version of the Code for Instructional Structure and Student Academic Response | Carta et al., (1988) |
| MS-CISSAR | Mainstream CISSAR | Greenwood et al., (2002) |
| MSIPCOT | Middle School Intervention Project Classroom Observation | Kennedy, (2014) |
| N/A | "Coding System" | Cobb, (1972) |
| N/A | "Coding System | Perkins, (1965) |
| OPTIC | Observing Pupils & Teachers In Classrooms | Merrett & Wheldall, (1986) |
| OPTIC | Observation Protocol for Technology Integration in the Classroom | Northwest Regional Educational Laboratory, (2014) |
| PACT | Performance Assessment for California Teachers | PACT Consortium, |

| | | |
|---|---|---|
| | | (2012) |
| PLATO | Protocol for Language Arts Teaching Observations | Institute for Research on Policy Education & Practice, (2011) |
| PROS | Positive Reinforcement Observation Schedule | Bersoff & Moyer, (1973) |
| REDSOCS | Revised Edition School Observation Coding System | Jacobs et al., (2000) |
| ROLE | Ramey Observation of Learning Essentials | Ramey & Ramey, (2002) |
| RTOP | Reformed Teaching Observation Protocol | MacIsaac & Falconer, (2002) |
| SCAN | Systematic Classroom Analysis Notation | Beeby et al., (1980) |
| SECOS | State-Event Classroom Observation System | Saidargas, (1997) |
| SECOS-R | State-Event Classroom Observation System-Research Edition | Saudargas & Fellers, (1986) |
| SGID | Small Group Instructional Diagnosis | Clark & Redmond, (1982) |
| SOCS | School Observation Coding System | McNeil et al., (1991) |
| SOS | Student Observation System | Reynolds & Kamphaus, (2004) |
| SOS | Stallings Observational System | Stallings & Needles, 1985 |
| SRSS | The Student Risk Screening Scale | Drummond, (1994) |
| STIR | Science Teacher Inquiry Rubric | Bodzin & Beerer, (2003) |
| STROBE | STROBE (not an acronym) | O'Malley et al., (2003) |
| TACL-PBS | Tool for Assessing Classroom Level-Positive Behavior Support | Ern, (2006) |
| TIES | Instructional Engagement Scale | Dickinson, (2008) |
| TOT | Time On Task | Anderson, (1975) |
| TPOT | The Teaching Pyramid Observation Tool | Snyder et al., (2013) |
| TRU Math | Teaching for Robust Understanding of Mathematics | Schoenfeld, (2013) |
| Uteach | UTeach Teacher Observation Protocol | Marder & Walkington, (2012) |
| UTOP | UTeach Observation Protocol | Walkington et al., (2012) |
| VOS | Vanth Observation System | Cox & Cordray, (2008) |
| YANKEES | Youth Assessment of Needs for Kids exhibiting Emotional problems in School | Nock & Kurtz's (2005)* renaming of REDSOCS |

# References:

Abikoff, H., & Gittelman, R. (1985). Classroom observation code-a modification of the Stony-Brook code. *Psychopharmacology Bulletin*, 21(4), 901-909.

Achenbach, T. M., (1983). Manual for the child behavior checklist and revised child behavior profile. Department of Psychiatry of the University of Vermont.

Achenbach, T. M. (1986). The direct observation form of the child behavior checklist (rev. ed.). Burlington, VT: University of Vermont, Department of Psychiatry.

Adrian, M., Zeman, J., & Veits, G. (2011). Methodological implications of the affect revolution: A 35-year review of emotion regulation assessment in children. *Journal of experimental child psychology*, *110*(2), 171-197.

Afzal, S., & Robinson, P. (2011). Natural affect data: Collection and annotation. In *New Perspectives on Affect and Learning Technologies* (pp. 55-70). Springer New York.

Aghababyan, A. (2014). E3: Emotions, Engagement and Educational Games. *Educational Data Mining*.

Aghababyan, A., Martin, T., & Harris-Brasiel, S. Understanding How Frustration and Confusion Manifest in Educational Games.

Anderson, L. W. (1981). Instruction & Time-on-Task: a Review. *Journal of Curriculum Studies*, 13(4), 289-303.

Appleton , J. J. Christenson , S. L. Kim , D. Reschly , A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44, 427-445.

Anderson, L. W. (1975). A Measure of Student Involvement in Learning: Time-on-Task. Report accessed at http://files.eric.ed.gov/fulltext/ED110504.pdf.

Andres, J. M. L., Rodrigo, M. M. T., Sugay, J. O., Baker, R. S., Paquette, L., Shute, V. J., Ventura, M., & Small, M. (2014) An Exploratory Analysis of Confusion Among Students Using Newton's Playground. *22nd International Conference on Computers in Education*.

Andres, J. M. L., & Rodrigo, M. M. T. (2014). Learning and Affect Trajectories Within Newton's Playground. *3rd International Workshop on ICT Trends in Emerging Economies.* Nara, Japan.

Atwater, J. B., Lee, Y., Montagna, D., Reynolds, L. H., & Tapia, Y. (2009). Classroom CIRCLE: Classroom Code for Interactive Recording of Children's Learning Environments. EBASS-Mobile: Ecobehavioral Assessment System Software. Kansas City, KS: Juniper Gardens Children's Project, Institute for Life Span Studies, University of Kansas.

Bakeman, R. (2000). Behavioral observation and coding. *Handbook of research methods in social and personality psychology.* Cambridge University Press, New York, 138-159.

Baker, J. A., Clark, T. P., Maier, K. S., & Viger, S. (2008). The differential influence of instructional context on the academic engagement of students with behavior problems. *Teaching and Teacher Education*, 24(7), 1876-1883.

Baker, R.S.J.d. (2007) Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. *Proceedings of ACM CHI 2007: Computer-Human Interaction*, 1059-1068.

Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004) Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, 531-540.

Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. (2004) Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". *Proceedings of ACM*

*CHI 2004: Computer-Human Interaction*, 383-390.

Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. (2008) Developing a Generalizable Detector of When Students Game the System. *User Modeling and User-Adapted Interaction*, 18, 3, 287-314.

Baker, R.S., DeFalco, J.A., Ocumpaugh, J., Paquette, L. (2014) Towards Detection of Engagement and Affect in a Simulation-based Combat Medic Training Environment. Paper presented at *2nd Annual GIFT User Symposium (GIFTSym2)*.

Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V. Kusbit, G., Ocumpaugh, J., Rossi, L. (2012) Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. *Proceedings of the 5th International Conference on Educational Data Mining*, 126-133.

Baker, R.S., Ocumpaugh, J. (2014) Cost-Effective, Actionable Engagement Detection at Scale. *Proceedings of the 7th International Conference on Educational Data Mining*, 345-346.

Baker, R.S.J.d., Ocumpaugh, J. (2015) Interaction-Based Affect Detection in Educational Software. In R.A. Calvo, S.K. D'Mello, J. Gratch, A. Kappas (Eds.), *Handbook of Affective Computing*. Oxford, UK: Oxford University Press.

Baker, R.S.J.d., Ocumpaugh, J.L., Gowda, S.M., Gowda, S.M., Heffernan, N.T. (2013) Ensuring Reliability of Educational Data Mining Detectors for Diverse Populations of Learners. Presentation at *CREA: Center for Culturally Responsive Evaluation and Assessment: Inaugural Conference*.

Baker, R.S., Ocumpaugh, J., Gowda, S.M., Kamarainen, A., Metcalf, S.J. (2014) Extending Log-Based Affect Detection to a Multi-User Virtual Environment for Science. To appear in *Proceedings of the 22nd Conference on User Modelling, Adaptation, and Personalization*.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C. (2010) Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. *International Journal of Human-Computer Studies*, 68 (4), 223-241.

Baker, R.S.J.d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., Yaron, D. (2011) The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction*.

Baker, R.S.J.d., Rodrigo, M.M.T., Xolocotzin, U.E. (2007) The Dynamics of Affective Transitions in Simulation Problem-Solving Environments. *Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*.

Baker, R. S., & Rossi, L. M. (2013). Assessing the Disengaged Behaviors of Learners. *Design Recommendations for Intelligent Tutoring Systems*, *1*, 155-166.

Baldwin, C. P. (1965). Naturalistic studies of classroom learning. *Review of Educational Research*, 107-113.

Bailey, W., Nowicki, S., & Cole, S. P. (1998). The ability to decode nonverbal information in African American, African and Afro-Caribbean, and European American adults. *Journal of Black Psychology*, *24*(4), 418-431.

Baron-Cohen, S. B., Golan, O., Wheelwright, S., & Hill, J. (2004). Mind reading: The interactive guide to emotions. London: Jessica Kingsley.

Braden, J. P. (2003). Psychological assessment in school settings. *Handbook of psychology*.

Beaupré, M.G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology*, *36*(3), 355-370.

Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin*, *32*, 16–26.

Beck, J., Rodrigo, M.M.T. (2014). Understanding Wheel Spinning in the Context of Affective Factors. *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*, 162-167.

Beeby, T., Burkhardt, H., & Caddy, R. (1980). SCAN: Systematic classroom analysis notation for mathematics lessons. Nottingham: England Shell Centre for Mathematics Education.

Beesley, A., Clark, T., Barker, J., Germeroth, C., & Apthorp, H. (2010). Expeditionary Learning Schools: Theory of Action and Literature Review of Motivation, Character, and Engagement. *Mid-continent Research for Education and Learning* (McREL).

Berry, D. J., Bridges, L. J., & Zaslow, M. J. (2004). Early Childhood Measures Profiles. *Child Trends' Report to the SEED Consortium of Federal Agencies* (including the Office of the Assistant Secretary for Planning & Evaluation of the U.S. Department of Health & Human Services, the National Institute of Child Health & Human Development, and the Office of Planning, Research & Evaluation of the Administration for Children & Families of the U.S. Department of Health &Human Services). Washington, DC.

Bersoff, D. N., & Moyer, D. (1973). Positive reinforcement observation schedule (PROS): Development and use. *American Psychological Association*, Montreal.

Biddle, B. J. (1967). Methods and concepts in classroom research. *Review of Educational Research*, 337-357.

Bieg, M., Goetz, T., & Lipnevich, A. A. (2014). What students think they feel differs from what they really feel–academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PloS one*, *9*(3), e92563.

Blanchard, E. G. (2014). Socio-Cultural Imbalances in AIED Research: Investigations, Implications and Opportunities. *International Journal of Artificial Intelligence in Education*, 1-25.

Bodzin, A. M., & Beerer, K. M. (2003). Promoting Inquiry-Based Science Instruction: The Validation of the Science Teacher Inquiry Rubric (STIR). *Journal of Elementary Science Education*, 15(2), 39-49.

Boekaerts, M. (2007). Understanding students' affective processes in the classroom. *Emotion in education*, 37-56.

Boller, K., & Sprachman, S. (1998). The Child-Caregiver Observation System. Instructor's Manual. Mathematica Policy Research.

Borich, G. D., Malitz, D., & Kugle, C. L. (1978). Convergent and discriminant validity of five classroom observation systems: Testing a model. *Journal of Educational Psychology*, *70*(2), 119.

Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology*, *11*(3), 263-280

Bowes, A. S., & Banilower, E. R. (2004). LSC classroom observation study: An analysis of data collected between 1997 and 2003. *Chapel Hill, NC: Horizon Research Inc*.

Briesch, A. M., & Volpe, R. J. (2007). Important considerations in the selection of progress-monitoring measures for classroom behaviors. *School Psychology Forum* 1, 59-74.

Brittain, K. (2012). M. Cl. Sc.(SLP) Candidate University of Western Ontario: School of Communication Sciences and Disorders This critical review examines whether speech-generating devices are effective in teaching children with autism new communicative

skills. Seven studies are reviewed here totalling.

Burry-Stock, J. A., & Oxford, R. L. (1994). Expert science teaching educational evaluation model (ESTEEM): Measuring excellence in science teaching for professional development. *Journal of Personnel Evaluation in Education*, 8(3), 267-297.

Calkins, D., Borich, G. D., Pascone, M., Kugle, C. L., & Marston, P. T. (1977). Generalizability of teacher behaviors across classroom observation systems. *The Journal of Classroom Interaction*, 9-22.

Carlson, S., Heimlich, J. E., & Storksdieck, M. (2011). Validating an Environmental Education Field Day Observation Tool. *International Electronic J. of Environmental Education*, *1*(3).

Carta, J. J., Greenwood, R., Schulte, D., Arreaga- Mayer, & Terry, B. (1988). *Code for Instructional Structure and Student Academic Response: Mainstream version* (MS-CISSAR). Kansas City, KS: University of Kansas, Bureau o f Child Research, Juniper Gardens Children's Project.

Casabianca, J. M., Lockwood, J. R., & McCaffrey, D. F. (2014). Trends in classroom observation scores. *Educational and Psychological Measurement*, 0013164414539163.

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*, *26*(3), 139-146.

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, 48(3), 219-246.

Christ, T. J., & Vining, O. (2006). Curriculum-based measurement procedures to develop multiple-skill mathematics computation probes: Evaluation of random and stratified stimulus-set arrangements. *School Psychology Review*, 35(3), 387.

Christ, T. J., Riley-Tillman, T. C., Chafouleas, S., & Jaffery, R. (2011). Direct behavior rating: An evaluation of alternate definitions to assess classroom behaviors. *School Psychology Review*, 40(2), 181.

Cicchetti, D.V. (1994). Guidelines, criteria, & rules of thumb for evaluating normed & standardized assessment instruments in psychology. *Psychological Assessment*, *6,*284–290.

Clark, D. J., & Redmond, M. V. (1982). Small Group Instructional Diagnosis: Final Report.

Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology*, 63(1), 74.

Cocea, M., Hershkovitz, A., Baker, R.S.J.d. (2009). The Impact of Off-Task and Gaming Behaviors on Learning: Immediate or Aggregate? *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 507-514.

Cone, J. D. (1978). The behavioral assessment grid (BAG): A conceptual framework and a taxonomy. *Behavior Therapy*, 9(5), 882-888.

Conners, C. K. (1997). Conners' Teacher Rating Scale--Revised (L). North Tonawanda, NY: Multi-Health Systems.

Cox, M. F., & Cordray, D. S. (2008). Assessing pedagogy in bioengineering classrooms: Quantifying elements of the "How People Learn" model using the VaNTH Observation System (VOS). *Journal of Engineering Education*, 97(4), 413-431.

Crosby, E. G., & French, J. L. (2002). Psychometric data for teacher judgments regarding the learning behaviors of primary grade children. *Psychology in the Schools*, 39(3), 235-244.

Csikszentmihalyi, M. (1990). *Flow: the Psychology of Optimal Experience*. Harper-Row, New York.

Dagami, M.M.C., Guo, B., Pating, M.B., Guia, T.F.G., Leonor, W.B.S., & Rodrigo, M.M.T. (2011). Determining the precursors of boredom. *Philippine Computing Journal*, *6*(1), 3-66.

Dancy, M. H., & Beichner, R. J. (2002). But are they learning? Getting started in classroom evaluation. *Cell Biology Education*, *1*(3), 87-94.

Danielson, C. (2011) The framework for teaching evaluation instrument, 2011 Edition. Accessed from: http://www.danielsongroup.org/article.aspx?page=FfTEvaluationInstrument.

Darr, C. W. (2012). Measuring student engagement: The development of a scale for formative use. *Handbook of Research on Student Engagement*. Springer US. 707-723.

DeFalco, J.A., Baker, R.S.J.d. (2013) Detection and Transition Analysis of Engagement and Affect in a Simulation-based Combat Medic Training Environment. *AIED 2013 Workshop on GIFT*.

Dettmers, S., Trautwein, U., Lüdtke, O., Goetz, T., Frenzel, A.C., & Pekrun, R. (2011). Students' emotions during homework in mathematics: Testing a theoretical model of antecedents & achievement outcomes. *Contemporary Educational Psychology*, *36*(1), 25-35.

Dewey, M. E. (1983). Coefficients of agreement. *British Journal of Psychiatry*, *143*(5), 487-489.

Dickinson, D. (2008). *Teacher Instructional Engagement Scale*. Nashville, TN: Vanderbilt.

Di Eugenio, B., & Glass, M. (2004). The kappa statistic: A second look. *Computational linguistics*, *30*(1), 95-101.

Dirr, P. J. (2003). Classroom observation protocols: Potential tools for measuring the impact of technology in the classroom. *Policy and Planning Series*, (104).

D'Mello, S.K., Graesser, A., Picard, R.W. (2007) Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22* (4), 53-61.

D'Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., Brawner, K. (2014) I Feel Your Pain: A Selective Review of Affect-Sensitive Instructional Strategies. In Sottilare, R., Graesser, A., Hu, X., & Goldberg, B. (Eds.). *Design Recommendations for Intelligent Tutoring Systems: Vol. 2 - Instructional Management*. Orlando, FL: U.S. Army Research Laboratory. 35-48.

Doctoroff, G. L., & Arnold, D. H. (2004). Parent-rated externalizing behavior in preschoolers: The predictive utility of structured interviews, teacher reports, and classroom observations. *Journal of Clinical Child and Adolescent Psychology*, *33*(4), 813-818.

Dombrowski, S. C., & Gischlar, K. L. (2015). Observing the Child. In *Psychoeducational Assessment and Report Writing* (pp. 43-62). Springer New York.

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. Journal of Youth and Adolescence, 40(12), 1649-1660.

Dowdy, E., Chin, J. K., & Quirk, M. P. (2013). Preschool Screening An Examination of the Behavioral and Emotional Screening System Preschool Teacher Form (BESS Preschool). *Journal of Psychoeducational Assessment*, *31*(6), 578-584.

Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. Early Childhood Research Quarterly, 25(1), 1-16.

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher*, *38*(7), 518-521.

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS).* Grants Pass, OR: Josephine County Mental Health Program.

Duvall, S. F., Jain, S., & Boone, D. (2010). An Observational Case Study of Four Second Grade General Education Students' Academic Responding and Inappropriate Behavior in the

Presence of a Disruptive Student with Disabilities. *Journal of Instructional Psychology*, *37*(4), 308.

Eastwood, J.D., Frischen, A., Fenske, M.J., & Smilek, D. (2012). The Unengaged Mind: Defining Boredom in Terms of Attention. *Perspectives on Psychological Science*, *7*(5), 482-495.

Eckert, T. L., & Lovett, B. J. (2013). Principles of Behavioral Assessment. *The Oxford Handbook of Child Psychological Assessment*, 366.

Elfenbein, H.A. (2006). Learning in emotion judgments: Training and the cross-cultural understanding of facial expressions. *Journal of Nonverbal Behavior*, *30*, 21–36.

Elfenbein, H.A., & Ambady, N. (2002a). Is there an in-group advantage in emotion recognition? *Psychological Bulletin*, *128*, 243–249.

Elfenbein, H.A., & Ambady, N. (2002b). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–235.

Elfenbein, H.A., & Ambady, N. (2003a). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality & Social Psychology*, *85*, 276–290.

Elfenbein, H.A., & Ambady, N. (2003b). Cultural Similarity's Consequences A Distance Perspective on Cross-Cultural Differences in Emotion Recognition. *Journal of Cross-Cultural Psychology*, *34*(1), 92-110.

Elfenbein, H.A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131.

Elfenbein, H.A., Mandal, M.K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: highlighting design and analytical techniques. *Emotion*, *2*(1), 75.

Elfenbein, H.A, Mandal, M., Ambady, N., Harizuka, S., & Kumar, S. (2004). Hemifacial differences in the in‐group advantage in emotion recognition. *Cognition & Emotion*, *18*(5), 613-629.

Ern, G.S. (2006). An examination of the relationship between the presence of critical components of classroom positive behavior support and student behavior.

Evertson, C. M., & Burry, J. A. (1988). Capturing Classroom Context: The Observation System as Lens for Assessment. *Journal of Personnel Evaluation in Education*, *2*(4), 297-320.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal*, *17*(1), 43-60.

Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, *2*, 212-236.

Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., & Eastwood, J. D. (2011). Development and validation of the multidimensional state boredom scale. *Assessment*, 1073191111421303.

Fancsali, S. E. Causal Discovery with Models: Behavior, Affect, and Learning in Cognitive Tutor Algebra.

Fancsali, S. E. (2013). Data-driven causal modeling of "gaming the system" and off-task behavior in Cognitive Tutor Algebra. In *NIPS Workshop on Data Driven Education*.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness--the development and correlates of a new scale. *Journal of Personality Assessment,* 50(1), 4-17.

Fisher, A., Godwin, K., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: when too much of a good thing may be bad. *Psychological*

*Science 25(7)*, 1362-1370.
Forness, S. R., & Guthrie, D. (1977). Stability of pupil behavior in short-term classroom observations. *Psychology in the Schools*.
Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. Regional Educational Laboratory Southeast.
Frick, T. W. (1990). Analysis of patterns in time: A method of recording and quantifying temporal relations in education. *American Educational Research Journal*, *27*(1), 180-204.
Gage, N. A., Prykanowski, D., & Hirn, R. (2014). Increasing Reliability of Direct Observation Measurement Approaches in Emotional and/or Behavioral Disorders Research Using Generalizability Theory. *Behavioral Disorders*, *39*(4).
Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review*, 3, 185–198.
Gee, J.P. (2007). Good video games + good learning: Collected essays on video games, learning, and literacy. Peter Lang Pub Incorporated: Bern, Switzerland.
Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–165.
Gibson, J., Hussain, J., Holsgrove, S., Adams, C., & Green, J. (2011). Quantifying peer interactions for research and clinical use: the Manchester Inventory for Playground Observation. *Research in developmental disabilities*, *32*(6), 2458-2466.
Gladman, M., & Lancaster, S. (2003). A review of the behaviour assessment system for children. *School Psychology International*, *24*(3), 276-291.
Godwin, K. E., Almeda, M. V., Petroccia, M., Baker, R. S., & Fisher, A. V. (2013). Classroom activities and off-task behavior in elementary school children. *Cognitive Science Society*.
Godwin, K,. Almeda, M., Skerbetz, M., Baker, R., & Fisher, A. (2014). Off-task behavior in elementary school children: examining changes in behavior across the school year and across instructional strategies. Learning and Instruction.
Goodman, Y. M. (2014). Observing Children in the Classroom. *Making Sense of Learners Making Sense of Written Language: The Selected Works of Kenneth S. Goodman and Yetta M. Goodman*, 197.
Graesser, A.C., Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of ducational Psychology*. **95**, 524–536.
Greenwood, C.R., Carta, J.J., Kamps, D., Terry, B., & Delquadri, J. (1994). Development and validation of standard classroom observation systems for school practitioners: Eco-Behavioral Assessment Systems Software (EBASS). *Exceptional Children*, 61(2), 197-210.
Greenwood, C. R., Delquadri, J., & Hall, R. V. (1978). Code for instructional structure and student academic response: CISSAR. Kansas City, KS: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.
Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*.
Grinder, E. (2007). Review of early childhood classroom observation measures. *Pennsylvania*

*Early Learning Standards–Classroom Observation Measures*, 3-10.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The Test Matters The Relationship Between Classroom Observation Scores and Teacher Value Added on Multiple Types of Assessment. *Educational Researcher*, 0013189X14544542.

Gu, H., Lai, S. L., & Ye, R. (2011). A cross-cultural study of student problem behaviors in middle schools. *School Psychology International*, *32*(1), 20-34.

Guia, T. F. G., Rodrigo, M. M. T., Dagami, M. M., Sugay, J. O., Macam, F. J. P., & Mitrovic, A. (2013). An Exploratory Study of Factors Indicative of Affective States of Students using SQL-Tutor. *Research & Practice on Technology Enhanced Learning*, 8(3), 411-430.

Guia, T. F. G., Rodrigo, M. M. T., Dagami, M. M. C, Sugay, J. O., Macam, F. J. P., & Mtriovic, A. (2011). Transitions of affective states in an intelligent tutoring system. *Philippine Computing Journal*, Dedicated Issue on Affect and Empathic Computing, 6(2), 31-35.

Hamre, B. K., Pianta, R. C., & Chomat-Mooney, L. (2009). Conducting classroom observations in school-based research.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 US early childhood and elementary classrooms. *Downloaded on March*, *27*, 2013.

Harms, T., Clifford, R. M., & Cryer, D. (1998). The Early Childhood Environment Rating sale (Rev. ed.). New York: Teachers College Press.

Harrison, P., & Oakland, T. (2003). Adaptive Behavior Assessment System (ABAS-II). San Antonio, TX: The Psychological Corporation.

Harniss, M. K., Caros, J., & Gersten, R. (2007). Impact of the design of US history textbooks on content acquisition and academic engagement of special education students: An experimental investigation. *Journal of Learning Disabilities*, 40(2), 100-110.

Hart, S. R., Stewart, K., & Jimerson, S. R. (2011). The student engagement in schools questionnaire (SESQ) and the teacher engagement report form-new (TERF-N): Examining the preliminary evidence. *Contemporary School Psychology*, 15(1), 67-79.

Hemmeter, M. L. (2001). *Assessment of Practices in Early Elementary Classrooms: APEEC*. New York: Teachers College Press.

Hennick, L. O. (2006). *Affective and behavioral parental constructs of adolescent behaviors* (Doctoral dissertation, University of Georgia).

Herba, C. M., Benson, P., Landau, S., Russell, T., Goodwin, C., Lemche, E., & Phillips, M. (2008). Impact of familiarity upon children's developing facial expression recognition. *Journal of Child Psychology and Psychiatry*, *49*(2), 201-210.

Herriott, R. E., & Firestone, W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher*, 14-19.

Hershkovitz, A., Baker, R.S.J.d., Gobert, J., Kauffman-Rogoff, Z., Wixon, M. (2012) Student Attributes, Affective States, and Engagement in Science Inquiry Microworlds. *The European Association for Research on Learning and Instruction (EARLI) SIG 20 Conference*.

Hershkovitz, A., Baker, R.S.J.d., Gobert, J., Nakama, A. (2012) A Data-driven Path Model of Student Attributes, Affect, and Engagement in a Computer-based Science Inquiry Microworld. *Proceedings of the International Conference on the Learning Sciences*.

Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review*, *33*(2), 258.

Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. *Best Practices in School Psychology*, 4, 993-1006.

Hoffman, J. V., Sailors, M., Duffy, G. R., & Beretvas, S. N. (2004). The effective elementary classroom literacy environment: Examining the validity of the TEX-IN3 observation system. *Journal of Literacy Research*, *36*(3), 303-334.

Hoge, R. D. (1985). The validity of direct observation measures of pupil classroom behavior. *Review of Educational Research*, 55(4), 469-483.

Hops, H., Davis, B., & Longoria, N. (1995). Methodological issues in direct observation: Illustrations with the Living in Familial Environments (LIFE) coding system. *Journal of Clinical Child Psychology, 24(2),* 193-203.

Hoque, M. E., & Picard, R. W. (2011, March). Acted vs. natural frustration and delight: Many people smile in natural frustration. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 354-359). IEEE.

Hurwitz, R. F. (1979). The reliability and validity of descriptive-analytic systems for studying classroom behaviours. *Focus on teaching: Readings in the observation and conceptualisation of teaching*, 111-118.

Hymavathy, C., Krishnamani, V., & Sumathi, c. (2014). Analyzing Learner Engagement to Enhance the Teaching-Learning Experience. *Proceedings of the 2nd IEEE International Conference on MOOCs, Innovation, and Technology in Education*.

Hyson, M.C., Hirsh-Pasek, K., & Rescorla, L. (1990). The classroom practices inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4-and 5-year-old children. *Early Childhood Research Quarterly*, 5(4), 475-494.

Indiana University Center for Postsecondary Research (IUCPR). (2003). *National survey of student engagement: The college student.*

Institute for Research on Policy Education & Practice. (2011). PLATO (Protocol for Language Arts Teaching Observations). Stanford, CA: Institute for Research on Policy Education & Practice.

Izard, C. E., King, K. A., & Finlon, K. J. (2007). Behavior and emotion expression observation system. University of Delaware.

Jacobs, J.R., Boggs, S.R., Eyberg, S.M., Edwards, D., Durning, P., Querido, J.G., & Funderburk, B.W. (2000). Psychometric properties and reference point data for the Revised Edition of the School Observation Coding System. *Behavior Therapy*, 31(4), 695-712.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*,*109*(19), 7241-7244.

Johnson, S.M., & Bolstad, O.D. (1983). Methodological issues in naturalistic observation: Some problems & solutions for field research. In L.A. Hamerlynck, L.C. Handy, & E.J. Mash (Eds.), *Behavior change: Methodology, Concepts & Practice*, Champaign, IL.: Research P.

Junker, B., Matsumura, L. C., Crosson, A., Wolf, M. K., Levison, A., Weisberg, Y., & Resnick, L. (2004). Overview of the Instructional Quality Assessment. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Junod, R. E. V., DuPaul, G. J., Jitendra, A. K., Volpe, R. J., & Cleary, K. S. (2006). Classroom observations of students with and without ADHD: Differences across types of engagement. *Journal of School Psychology*, *44*(2), 87-104.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavior and Emotional Screening*

*System (BASC-2 BESS)*. San Antonio, TX: Pearson.
Karweit, N. (1982). Time on Task: A Research Review. Report No. 332.
Karweit, N., Slavin, R.E. (1982) Time-On-Task: Issues of Timing, Sampling, and Definition. *Journal of Experimental Psychology,* 74 (6), 844-851.
Kassam, K.S., & Mendes, W.B. (2013). The effects of measuring emotion: Physiological reactions to emotional situations depend on whether someone is asking. *PLOS One, 8(6),* 649-659.
Kennedy, P. (2014). Measuring the Effects of Instructional Environment and Student Engagement on Reading Achievement for Struggling Readers in Middle School.
Kerr, D. M., Kent, L., & Lam, T. C. (1985). Measuring program implementation with a classroom observation instrument: The Interactive Teaching Map. *Evaluation Review*, 9(4), 461-482.
Khachatryan, G. A., Romashov, A. V., Khachatryan, A. R., Gaudino, S. J., Khachatryan, J. M., Guarian, K. R., & Yufa, N. V. (2014). Reasoning Mind Genie 2: An Intelligent Tutoring System as a Vehicle for International Transfer of Instructional Methods in Mathematics. *International Journal of Artificial Intelligence in Education*, *24*(3), 333-382.
Kilbride, J. E., & Yarczower, M. (1983). Ethnic bias in the recognition of facial expressions. *Journal of Nonverbal Behavior*, *8*(1), 27-41.
King, J. (2013). Silence in the second language classrooms of Japanese universities. *Applied Linguistics*, 34(3), 325-343.
Kisker, E. E., Boller, K., Nagatoshi, C., Sciarrino, C., Jethwani, V., Zavitsky, T., Ford, M., & Love, J. M. (2003). Resources for Measuring Services and Outcomes in Head Start Programs Serving Infants and Toddlers. Washington, D.C.: Prep. by Mathematica Policy Research, Inc. for the Office of Planning, Research & Evaluation of the Administration for Children & Families of the U.S. Department of Health & Human Services.
Kort, B., Reilly, R., Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion. *Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges*, IEEE Computer Society, Madison, Wisconsin. 43–48.
Kravcik, M., Santos, O. C., & Boticario, J. G. (2014) 4th International Workshop on Personalization Approaches in Learning Environments (PALE 2014).
Krumhuber, E. G., Kappas, A., & Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: a review. *Emotion Review*, *5*(1), 41-46.
Lagud, M. C. V. & Rodrigo, M. M. T. (2010). The affective and learning profiles of students while using an intelligent tutoring system for algebra. In V. Aleven, J. Kay & J. Mostow (Eds.), *Proceedings of Intelligent Tutoring Systems. Lecture Notes in Computer Science*, Vol. 6094, 255-263.
Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth grade classrooms. *Journal of Educational Psychology*, 59, 320-324.
Lamb, M. L., & Swick, K. J. (1975). A Historical Overview of Classroom Teacher Observation. In *The Educational Forum,* 39(2), 238-247.
La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 409-426.
Lawrenz, F., Huffman, D., Appeldoorn, K., & Sun, T. (2002). Classroom observation handbook. *CETP Core Evaluation—Classroom Observation Protocol. Minneapolis: Center for Applied Research in Educational Improvement, College of Education and Human*

*Development, University of Minnesota*.
Lee, D. M., Rodrigo, M. M. T. R., Baker, Ryan S. J. D., Sugay, J. O., & Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. In S. D'Mello & A Graesser (Eds.): ACII 2011, Part I, LNCS 6974, 175-184, Berlin Heidelberg: Springer-Verlag.
Leff, S. S., & Lakin, R. (2005). Playground-based observational systems: A review and implications for practitioners and researchers. *School Psychology Review*, *34*(4), 475.
Leff, S. S., Thomas, D. E., Shapiro, E. S., Paskewich, B., Wilson, K., Necowitz-Hoffman, B., & Jawad, A. F. (2011). Developing and validating a new classroom climate observation assessment tool. *Journal of School Violence*, 10(2), 165-184.
Lehman, B., Matthews, M., D'Mello, S., & Person, N. (2008). What are you feeling? Investigating student affective states during expert human tutoring sessions. In *Intelligent Tutoring Systems* (pp. 50-59). Springer Berlin Heidelberg.
Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2), 112-129.
Lett, N. J., & Kamphaus, R. W. (1992). Validation of the BASC Teacher Rating Scale by the BASC Student Observation Scale. *Canadian Journal of School Psychology*, *13*(1), 1-14.
Linnenbrink-Garcia, L., & Pekrun, R. (2011a). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1-3.
Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. (2011b). Affect and engagement during small group instruction. *Contemporary Educational Psychology*, *36*(1), 13-24.
Lindquist, K. A., & Gendron, M. (2013). What's in a Word? Language Constructs Emotion Perception. *Emotion Review*, *5*(1), 66-71.
Lippman, L., & Rivers, A. (2008). Assessing school engagement: A guide for out-of-school time program practitioners. *A Research-to-Results brief.* Washington, DC: Child Trends.
Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R.S.J.d. (2013) Sequences of Frustration and Confusion, and Learning. *Proceedings of the 6th International Conference on Educational Data Mining*, 114-120.
Lloyd, J.W., Loper, A.B. (1986) Measurement & Evaluation of Task-Related Learning Behavior: Attention to Task and Metacognition. *School Psychology Review,* 15 (3), 336-345.
MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via RTOP. *The Physics Teacher*, 40(8), 479-485.
Malia, J. A. (2007). A reader's guide to family stress literature. *Journal of Loss and Trauma*, *12*(3), 223-243.
Mandal, M. K. (1996). Similarities and variations in facial expressions of emotions: Cross-cultural evidence. *International Journal of Psychology*, *31*(1), 49-58.
Mandal, M. K., Asthana, H. S., Madan, S. K., & Pandey, R. (1992). Hemifacial display of emotion in the resting state. *Behavioural Neurology*, *5*(3), 169-171.
Mandal, M. K., Harizuka, S., Bhushan, B., & Mishra, R. C. (2001). Cultural variation in hemifacial asymmetry of emotion expressions. *British Journal of Social Psychology*, *40*(3), 385-398.
Marchand, G. C., & Gutierrez, A. P. (2012). The role of emotion in the learning process: Comparisons between online and face-to-face learning settings. *The Internet and Higher Education*, *15*(3), 150-160.
Marder, M., & Walkington, C. (2012). *UTeach Teacher Observation Protocol*. Accessed: https://

wikis.utexas.edu/pages/viewpageattachments.action?pageId=6884866&sortBy=date&highlight=UTOP_Physics_2009.doc.&.

Markham, R., & Wang, L. (1996). Recognition of emotion by Chinese and Australian children. *Journal of Cross-Cultural Psychology*, *27*(5), 616-643.

Martin, P. A., Daley, D., Hutchings, J., Jones, K., Eames, C., & Whitaker, C. J. (2010). The teacher-pupil observation tool (T-POT) development and testing of a New classroom observation measure. *School Psychology International*, *31*(3), 229-249.

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: comment on Elfenbein and Ambady (2002) and evidence. *Psychological Bulletin*, 128(2), 236-242.

Matsumoto, D., Hwang, H. S., & Yamada, H. (2010). Cultural differences in the relative contributions of face and context to judgments of emotions. *Journal of Cross-Cultural Psychology*. 1-21.

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*(2), 211-224.

McClenney, K., Marti, C. N., & Adkins, C. (2012). Student Engagement and Student Outcomes: Key Findings from" CCSSE" Validation Research. Community College Survey of Student Engagement.

McConaughy, S.H., & Achenbach, T.M. (2009). Manual for the ASEBA Direct Observation Form. Burlington: U. of Vermont, Research Center for Children, Youth, & Families.

McKinney, C., & Morse, M. (2012). Assessment of disruptive behavior disorders: Tools and recommendations. *Professional Psychology: Research & Practice*, *43*(6), 641.

McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*(3), 477-505.

McNeil, C.B., Eyberg, S., Eisenstadt, T.H., Newcomb, K., & Funderburk, B. (1991). Parent–Child Interaction Therapy with behavior problem children: Generalization of treatment effects to the school setting. *Journal of Clinical Child Psychology*, 20, 140–151.

Medley, D.M., & Mitzel, H.E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed*.), Handbook of Research on Teaching.* Chicago: Rand McNally.

Medley, D. M., & Norton, D. P. (1971). The concept of reliability as it applies to behavior records. Paper presented to the American Psychological Association, Washington, D.C.

Merrett, F., & Wheldall, K. (1986). Observing pupils and teachers in classrooms (OPTIC): A behavioural observation schedule for use in schools. *Educational Psychology,* 6(1), 57-70.

Merrell, K. W., Streeter, A. L., Boelter, E. W., Caldarella, P., & Gentry, A. (2001). Validity of the home and community social behavior scales: Comparisons with five behavior-rating scales. *Psychology in the Schools*, *38*(4), 313-325.

Metcalf, S., Kamarainen, A., Tutwiler, M. S., Grotzer, T., & Dede, C. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *3*(1), 86-90.

Miller, F. G., Chafouleas, S. M., Riley-Tillman, T. C., & Fabiano, G. A. (2014). Teacher Perceptions of the Usability of School-Based Behavior Assessments. *Behavioral Disorders,* 39(4).

Miller, W. L., Petsche, K., Baker, R. S., Labrum, M. J., & Wagner, A. Z. Boredom Across Activities, and Across the Year, within Reasoning Mind.

Miller, W.L., Baker, R., Labrum, M., Pestche, K., Liu, Y-H., Wagner, A. (in press) Automated

Detection of Proactive Remediation by Teachers in Reasoning Mind Classrooms. *Proceedings of the 4th Conference on Learning Analytics and Knowledge*.

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *J. Educ. Psychol.* **88**, 203–214.

Muijs, D., Chapman, C., & Armstrong, P. (2014). Acculturation or innovation? The pedagogical practices of teachers on an ambitious, alternative certification programme. *Learning Teaching from Experience: Multiple Perspectives and International Contexts*, 21.

Nesselrodt, P. S., & Schaffer, E. C. (1993, April). The ISERP programme: A revised classroom observation instrument. A*nnual meeting of the American Educational Research Association, Atlanta, GA*.

Ndirangu, M., Njumbi, J. N., & Sogomo, K. C. (2011). Effects of Using the Dimensions of Creativity in Teaching Mathematics on Students' Achievement and Perception of Classroom Learning Environment. *Journal of Technology & Education in Nigeria*, 16(1), 109-121.

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology*, *36*(1), 49-59.

Nickerson, A. B., & Fishman, C. (2009). Convergent and divergent validity of the Devereux Student Strengths Assessment. *School Psychology Quarterly*, *24*(1), 48.

Nock, M. K., & Kurtz, S. M. (2005). Direct behavioral observation in school settings: Bringing science to practice. Cognitive and Behavioral Practice, 12(3), 359-370.

Northwest Regional Educational Laboratory. (2004). *OPTIC—Observation protocol for technology integration in the classroom.* Portland, OR: Author.

Nowicki, S., Glanville, D., & Demertzis, A. (1998). A test of the ability to recognize emotion in the facial expressions of African American adults. *Journal of Black Psychology*, *24*(3), 335-350.

Nunn, R. (2011). Improving method-in-use through classroom observation. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*(1), 55-70.

O'Connor, E. A., & Fish, M. C. (1997). Differences between the Classrooms of Expert and Novice Teachers on the Dimensions of the "Classroom Systems Observation Scale."

Ocumpaugh, J., Baker, R.S.J.d., Gaudino, S., Labrum, M.J., Dezendorf, T. (2013) Field Observations of Engagement in Reasoning Mind. *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, 624-627.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., Heffernan, C. (2014) Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology, 45* (3), 487-501.

Ocumpaugh, J., Baker, R.S., Kamarainen, A.M., Metcalf, S.J. (2014) Modifying Field Observation Methods on the Fly: Metanarrative and Disgust in an Environmental MUVE. *Proceedings of PALE 2013: The 4th International Workshop on Personalization Approaches in Learning Environments*, 49-54.

Ocumpaugh, J., Baker, R.S.J.d., Rodrigo, M.M.T. (2012) *Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0*. Technical Report. New York, NY: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, 71, 137–149.

O'Malley, K. J., Moran, B. J., Haidet, P., Seidel, C.L., Schneider, V., Morgan, R.O., & Richards,

B. (2003). Validation of an observation instrument for measuring student engagement in health professions settings. *Evaluation & the Health Professions*, 26(1), 86-103.

Overton, T. (2004). Promoting academic success through environmental assessment. *Intervention in School & Clinic*, 39(3), 147-153.

PACT Consortium (2012) Performance Assessment for California Teachers. (2012) A brief overview of the PACT assessment system. Accessed: http://www.pacttpa.org/_main/hub.php?pageName=Home.

Pantic, M., & Rothkrantz, L. J. M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*(9), 1370–1390.

Pardos, Z.A., Baker, R.S.J.d., San Pedro, M.O.C.Z., Gowda, S.M., Gowda, S.M. (2013) Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, 117-124.

Paquette, L., de Carvalho, A.M.J.A., Baker, R.S., Ocumpaugh, J. (2014) Reengineering the Feature Distillation Process: A Case Study in the Detection of Gaming the System. *Proceedings of the 7th International Conference on Educational Data Mining*, 284-287.
Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014, January). Sensor-Free Affect Detection for a Simulation-Based Science Inquiry Learning Environment. In *Intelligent Tutoring Systems* (pp. 1-10). Springer International Publishing.

Perkins, H. V. Classroom behavior and underachievement. *American Educational Research Journal,* 1965, 2, 1-12.

Perreault Jr, W. D., & Leigh, L. E. (1989). Reliability of nominal data based on qualitative judgments. *Journal of Marketing Research (JMR)*, *26*(2).

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282). Springer US.

Pianta, R., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System*. Baltimore: Paul H. Brookes.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I., & Judson, E. (2000). Reformed teaching observation protocol (RTOP) reference manual. *Tempe, Arizona: Arizona Collaborative for Excellence in the Preparation of Teachers*.

Planalp, S., DeFrancisco, V.L., Rutherford, D. (1996) Varieties of Cues to Emotion in Naturally Occurring Settings. *Cognition and Emotion, 10* (2), 137-153.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental psychology*, *45*(3), 605.

Porayska-Pomsta, K., Mavrikis, M., D'Mello, S., Conati, C., & Baker, R. S. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education*, *22*(3), 107-140.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement & personality. *Research in Higher Education*, *46*(2), 127-152.

Raffaele Mendez, L. M., Hoy, B. D., Sundman-Wheat, A. N., & Cunningham, J. (2011). Research Advances in Understanding Emotional Dysregulation in Youth. *Communique*, *40*(3), 1-6.

Ramey, S. L. & Ramey, C. T. (2002). The Ramey Observation of Learning Essentials (ROLE), Washington, DC: Georgetown

Ramsay, M. C., Reynolds, C. R., & Kamphaus, R. W. (2002). *Essentials of behavioral assessment* (Vol. 37). John Wiley & Sons.

Reeve, J., & Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267.

Reschly, A. L., Betts, J., & Appleton, J. J. (2014). An examination of the validity of two measures of student engagement. *International Journal of School & Educational Psychology*, 2(2), 106-114.

Reynolds, C. R., & Kamphaus, R. W. (1998). *BASC: Behavior assessment system for children: Manual*. American Guidance Service.

Rhodes, G., & Lynskey, M. (1990). Face perception: Attributions, asymmetries and stereotypes. *British Journal of Social Psychology*, *29*(4), 375-377.

Riley-Tillman, T. C., Kalberer, S. M., & Chafouleas, S. M. (2005). Selecting the right tool for the job: A review of behavior monitoring tools used to assess student response to intervention. *The California School Psychologist, 10*(1), 81-91.

Ritchie, S., Howes, C., Kraft-Sayre, M. & Weiser, B. (2001). Emergent Academic Snapshot Scale. Los Angeles: UCLA.

Rivera, H. H., & Tharp, R. G. (2004). A Study of a Classroom Observation System. Observational research in US classrooms: New approaches for understanding cultural and linguistic diversity, 205.

Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B. (2007) Affect and Usage Choices in Simulation Problem Solving Environments. *Proceedings of Artificial Intelligence in Education 2007*, 145-152.

Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A. M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., & Viehland, N.J.B. (2007). Affect and usage choices in simulation problem-solving environments. In R. Luckin, K.R. Koedinger, J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, Amsterdam, The Netherlands: IOS Press. 145-152.

Rodrigo, M.M.T., Baker, R.S.J.d., d'Mello, S., Gonzalez, M.C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., Viehland, N.J.B. (2008) Comparing Learners' Affect While Using an Intelligent Tutoring Systems and a Simulation Problem Solving Game. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 40-49.

Rodrigo, M. M. T. & Baker, R. S. J. d. , & Rossi, L. (2013). Student Off-Task Behavior in Computer-Based Learning in the Philippines: Comparison to Prior Research in the USA. *Teachers College Record*, 115(10), 1-27.

Rodrigo, M.M.T., Baker, R.S.J.d. (2011) Comparing Learners' Affect While Using an Intelligent Tutor and an Educational Game. *Research & Practice in Technology Enhanced Learning*, 6 (1), 43-66.

Rodrigo, M. M. T. (2011). The dynamics of students' affective transitions while using a pre-algebra game. *Simulation & Gaming*, 42(1),85-99.

Rodrigo, M. M. T. & Baker, R. S. J. d. (2009). Coarse-grained detection of student frustration in an introductory programming course. *Proceedings of the International Computer Education Research Workshop*. Berkeley, CA, 75-80.

Rodrigo, M.M.T., Baker, R.S., Jadud, M.C., Amarra, A.C.M., Dy, T., Espejo-Lahoz, M.B.V., Lim, S.A.L., Pascua, S.A.M.S., Sugay, J.O., & Tabanao, E. (2009). Affective and

behavioral predictors of novice programmer achievement. *Innovation & Technology in Computer Science Education*, Universite Pierre et Marie Curie, Paris, France. ACM SIGCSE Bulletin, 41(3), 156-160.

Rodrigo, M.M.T., Rebolledo-Mendez, G., Baker, R.S.J.d., du Boulay, B., Sugay, J.O., Lim, S.A.L., Lahoz, M.B.E., Luckin, R. (2008). The effects of motivational modeling on affect in an intelligent tutoring system. In Chan, T.W., Biswas, G., Chen, F.C., Chen, S., Chou, C., Jacobson, M., Kinshuk, Klett, F., Looi, C.K., Mitrovic, T., Mizoguchi, R., Nakabayashi, K., Reimann, P., Suthers, D., Yang, S., & Yang, J.C. (Eds.). *Proceedings of the International Conference on Computers in Education*, 49-56.

Rodrigo, M. M. T., Anglo, E. A., Sugay, J., & Baker, R. (2008). Use of Unsupervised Clustering to Characterize Learner Behaviors and Affective States while Using an Intelligent Tutoring Systems. In Chan, T. W., Biswas, G., Chen, F.-C., Chen, S., Chou, C., Jacobson, M., Kinshuk, Klett, F., Looi, C.-K., Mitrovic, T., Mizoguchi, R., Nakabayashi, K., Reimann, P., Suthers, D., Yang, S., & Yang, J.-C. (Eds.). *International Conference on Computers in Education*, 57-64.

Rodrigo, M. M. T., Baker, R. S. J. d., D'Mello, S., Gonzalez, M. C. T., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sugay, J. O., Tep, S., & Viehland, N. J. B., (2008). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. In Woolf, B. P., Aimeur, E., Nkambou, R. & Lajoie, S. P. (Eds.) *Proceedings of Intelligent Tutoring Systems, 9th International Conference,* Montreal, Canada. Lecture Notes in Computer Science 5091, Springer, 40-49.

Rodrigo, M. M. T., Baker, R. S. J. d., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sevilla, L. R. S., Sugay, J. O., Tep, S., & Viehland, N. J. B. (2007). Affect and usage choices in simulation problem-solving environments. In R. Luckin, K. R. Koedinger, J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, Amsterdam: IOS Press. 145-152.

Rodrigo, M.M.T., Rebolledo-Mendez, G., Baker, R.S.J.d., du Boulay, B., Sugay, J.O., Lim, S.A.L., Espejo-Lahoz, M.B., Luckin, R. (2008). The Effects of Motivational Modeling on Affect in an Intelligent Tutoring System. *Proceedings of International Conference on Computers in Education*, 57-64.

Roehrig, A. D., & Christesen, E. (2010). Development and use of a tool for evaluating teacher effectiveness in grades K-12. In *Innovative Assessment for the 21st Century*. Springer: US. 207-228.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement a meta-analytic approach. *Review of Educational Research*, *81*(4), 493-529.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. Travers (Ed.), *Second Handbook of Research on Teaching.* Chicago, Ill.: Rand McNally.

Ross, S. M., & Smith, L. J. (1996). Classroom observation measure observer's manual. *Memphis, TN: University of Memphis/Memphis City Schools*.

Rush, S. C. (2013). Organisation, Interpretation, and Presentation of Classroom Observation Data: A Demonstration using Transana Qualitative Video Analysis Software. *Information Technology, Education & Society*, 14(1), 41-49.

Saklofske, D. H., Joyce, D. K., Sulkowski, M. L., & Climie, E. A. (2013). 15 Models for the Personality Assessment of Children and Adolescents. *The Oxford Handbook of Child Psychological Assessment*, 348.

Sadatsafavi, M., Najafzadeh, M., Lynd, L., & Marra, C. (2008). Reliability studies of diagnostic tests are not using enough observers for robust estimation of interobserver agreement: a simulation study. *Journal of Clinical Epidemiology*, *61*(7), 722-727.

Saginor, N., & Hyjek, P. (2005). Observing Standards-Based Classrooms: The Vermont Classroom Observation Tool (VCOT). *Montpelier, VT: Vermont Institutes*.

Sammons, P., & Ko, J. (2008). Using Systematic Classroom Observation Schedules to Investigate Effective Teaching: Overview of Quantitative Findings. An Effective Classroom Practice (ECP) Project Report. *Effective Classroom Practice (ECP) ESRC Project Report (RES-000-23-1564). Swindon: ESRC*.

Sarama, J., & Clements, D. H. (2007). Manual for Classroom Observation of Early Mathematics: Environment and Teaching (COEMET) V.3.

San Pedro, M.O.Z., Ocumpaugh, J.L., Baker, R.S., Heffernan, N.T. (2014) Predicting STEM and Non-STEM College Major Enrollment from Middle School Interaction with Mathematics Educational Software. *Proceedings of the 7th International Conference on Educational Data Mining*, 276-279.

San Pedro, M. O. Z., Baker, R. S., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *Proceedings of the 6th international conference on educational data mining* (pp. 177-184).

San Pedro, M.O.Z., Baker, R.S.J.d. & Rodrigo, M.M.T. (2014) Carelessness and Affect in an Intelligent Tutoring System for Mathematics. *International Journal of Artificial Intelligence in Education.*

San Pedro, M.O.C., Baker, R.S.J.d., Rodrigo, M.M. (2011) The Relationship between Carelessness and Affect in a Cognitive Tutor. *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction*.

San Pedro, M. O. C. Z., Baker, R. S. J D., & Rodrigo, M. M. T. (2011). Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics. In G. Biswas, S. Bull, J. Kay, & A Mitrovic. *Proceedings of the Conference on Artificial Intelligence in Education. Lecture notes in Computer Science* Vol. 6738. Springer Berlin/Heidelberg. 304-311.

Sanetti, L. M. H., & Collier-Meek, M. A. (2014). Increasing the Rigor of Procedural Fidelity Assessment: An Empirical Comparison of Direct Observation and Permanent Product Review Methods. *Journal of Behavioral Education*, 23(1), 60-88.

Saudargas, R. A. (1997). State-event classroom observation system (SECOS). Observation manual.

Saudargas, R. A., & Fellers, G. (1986). State-Event Classroom Observation System: Research edition (SECOS-R). Knoxville: University of Tennessee, Department of Psychology.

Sauter, D. A., & Eisner, F. (2013). Commonalities outweigh differences in the communication of emotions across human cultures. *Proceedings of the National Academy of Sciences*, *110*(3), E180-E180.

Sciarra, D. T., & Seirup, H. J. (2008). The multidimensionality of school engagement and math achievement among racial groups. *Professional School Counseling,* 11(4), 218-228.

Schimmack, U. (1996). Cultural Influences on the Recognition of Emotion by Facial Expressions Individualistic or Caucasian Cultures? *Journal of Cross-Cultural Psychology*, *27*(1), 37-50.

Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, *45*(4), 607-621.

Schultz, S., & Arroyo, I. (2014). Tracing Knowledge and Engagement in Parallel in an Intelligent Tutoring System. In *Proceedings of the 7th Annual International Conference on*

*Educational Data Mining*.
Sciarra, D. T., & Seirup, H. J. (2008). The multidimensionality of school engagement and math achievement among racial groups. *Professional School Counseling,* 11(4), 218-228.
Shapiro, E. S. (2003). *Behavioral Observation of Students in Schools (BOSS)*. Computer Software. San Antonio, TX: Psychological Corporation.
Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*(2), 158.
Schussler, D. L. (2009). Beyond content: How teachers manage classrooms to facilitate intellectual engagement for disengaged students. *Theory Into Practice*, 48(2), 114-121.
Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257-268.
Simon, A., & Boyer, E. G. (1970). Mirrors for Behavior, An Anthology of Classroom Observation Instruments, 1970 Supplement, Vols. A and B.
Skinner, C. H., Rhymer, K. N., & McDaniel, E. C. (2000). Naturalistic direct observation in educational settings. Conducting school-based assessments of child and adolescent behavior, 21-54.
Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2008). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*.c
Skinner, M., & Mullen, B. (1991). Facial asymmetry in emotional expression: A meta-analysis of research. *British Journal of Social Psychology*, *30*(2), 113-124.
Smith, W. M. (1998). Hemispheric and facial asymmetry: faces of academe. *Journal of cognitive neuroscience*, *10*(6), 663-667.
Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). Early Language & Literacy Classroom Observation Tool: K-3. Paul H. Brookes.
Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education*, *12*(4), 618-627.
Snyder, P. A., Hemmeter, M. L., Fox, L., Bishop, C. C., & Miller, M. D. (2013). Developing and gathering psychometric evidence for a fidelity instrument: The Teaching Pyramid Observation Tool–Pilot version. *Journal of Early Intervention*, 1053815113516794.
Soukakou, E. P. (2012). Measuring quality in inclusive preschool classrooms: Development and validation of the Inclusive Classroom Profile (ICP*). Early Childhood Research Quarterly,* 27(3), 478-488.
Spanjers, D. M., Burns, M. K., & Wagner, A. R. (2008). Systematic direct observation of time on task as a measure of student engagement. *Assessment for Effective Intervention*, 33(2), 120-126.
Spears, T. (2013). Promoting Classroom Engagement through instructional Practices Using the Common Core State Standards for Mathematics. Doctoral dissertation, Bowling Green State University.
Stallings, J. (1973). *Follow Through Classroom Observation Evaluation, 1971-72*. Stanford Research Institute.
Stallings, J., & Needles, M. (1985). *Stallings Observation Instrument (Revised Edition)*. Stanford, CA: SRI International.

Stanley, S.O., & Greenwood, C.R. (1981). *Code for instructional structure and student academic response (CISSAR): Observers' manual.* Kansas City, KS: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.

Steege, M. W., Davin, T., & Hathaway, M. (2001). Reliability and accuracy of a performance-based behavioral recording procedure. *School Psychology Review*.

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, *35*(2), 87-100.

Stringer, P., & May, P. (1981). Attributional asymmetries in the perception of moving, static, chimeric and hemisected faces. *Journal of Nonverbal Behavior*, *5*(4), 238-252.

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, 19(3), 375-397.

Stuessy, C. (2006). Mathematics and science classroom observation protocol system (MSCOPS): Classroom observation and videotape analysis of classroom learning environments. In *A manual prepared for a 2-day workshop prepared for mentors of intern teachers in the PLC-MAP project: College Station, TX, Department of Teaching, Learning, and Culture*.

Subkoviak, M. J., & Baker, F. B. (1977). Test theory. *Review of Research in Education*, 275-317.

Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing Quality in the Early Years-Early Childhood Environment Rating Scale Extension (ECERS-E): Four Curricular Subscales*. Trentham Books.

Tapp, J., Wehby, J. H., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers, 27,* 25-31.

Taylor, B. M., & Pearson, P. D. (2000). The CIERA school change classroom observation scheme. *Minneapolis: University of Minnesota*.

Thomas, L., & Jere E.. Brophy. (1994). Looking in Classrooms. New York: Harper & Row.

Tooth, L. R., & Ottenbacher, K. J. (2004). The κ statistic in rehabilitation research: An examination. *Archives of physical medicine and rehabilitation*, *85*(8), 1371-1376.

Towstopiat, O. (1984). A review of reliability procedures for measuring observer agreement. *Contemporary educational psychology*, *9*(4), 333-352.

Tze, V. M., Klassen, R. M., Daniels, L. M., Li, J. C. H., & Zhang, X. (2012). A cross-cultural validation of the Learning-Related Boredom Scale (LRBS) with Canadian and Chinese college students. *Journal of Psychoeducational Assessment*, 0734282912443670.

University of Michigan (2006). Learning mathematics for teaching. A coding rubric for measuring the mathematical quality of instruction (Technical Report LMT1.06). Ann Arbor, MI: University of Michigan, School of Education.

VanTassel-Baska, J. (2004). Assessing classroom practice: The use of a structured observation form. *Designing and utilizing evaluation for gifted program improvement*, 87-108.

Van Tilburg, W. A., & Igou, E. R. (2012). On boredom: Lack of challenge and meaning as distinct boredom experiences. *Motivation and Emotion*, *36*(2), 181-194.

Vaughn, S., & Briggs, K. L. (Eds.). (2003). *Reading in the classroom: Systems for the observation of teaching and learning*. PH Brookes Publishing Company.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360-363.

Volpe, R.J., DiPerna, J.., Hintze, J.M., & Shapiro, E.S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, *34*(4), 454.

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the Direct Observation Form. *School Psychology Review*, 38(3), 382.

Wagner, M., Sumi, W. C., Woodbridge, M. W., Javitz, H. S., & Thornton, S. P. (2009). Coordinating research and implementation of evidence-based school interventions for children with serious behavior problems. *Journal of Emotional and Behavioral Disorders*, 17, 244–249.

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., & Marder, M. (2012). Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program.

Wang, M. T., & Degol, J. (2014). Staying Engaged: Knowledge and Research Needs in Student Engagement. *Child Development Perspectives*, *8*(3), 137-143.

Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education*, *10*(4), 205-212.

Waxman, H. C., Wang, M. C., Lindvall, M., & Anderson, K. A. (1988). Classroom observation schedule technical manual. *Pittsburgh: University of Pittsburgh, Learning Research and Development Center*.

Weathers, M. D., Frank, E. M., & Spell, L. A. (2002). Differences in the communication of affect: Members of the same race versus members of a different race. *Journal of Black Psychology*, *28*(1), 66-77.

Weaver, D., Dick, T., Higgins, K., Marrongelle, K., Foreman, L., & Miller, N. (2005). OMLI classroom observation protocol. *Portland, OR: RMC Research Corporation*.

Weinholtz, D., Everett, G., Albanese, M. & Shymansky, J. (1986). The Attending Round Observation System: A proceedure for describing teaching during attending rounds. *Evaluation & the Health Professions*, 9, 75-89.

Weinrott, M. R., Jones, R. R., & Boler, G. R. (1981). Convergent and discriminant validity of five classroom observation systems: A secondary analysis. *Journal of Educational Psychology*, *73*(5), 671.

Whitcomb, S., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents*. Routledge.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, *8*(4), 367-372.

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive development*, *23*(2), 291-312.

Williams, B., & Carvalho, I. (2010). Using the LAMM classroom observation system to facilitate the adoption of active learning methodologies in engineering education. Proc of the Annual Conference of the European Society for Engineering Education (SEFI) 19-22.

Wilson, J. A., Spelman, B. J., & Trew, K. J. (1976). Experimental validation of two classroom observation systems. *Journal of Educational Psychology*, *68*(6), 742.

Wirtz, M., & Kutschmann, M. (2007). Analyzing interrater agreement for categorical data using Cohen's kappa and alternative coefficients. *Die Rehabilitation*, *46*(6), 370-377.

Wixon, M., & Arroyo, I. (2014). When the Question is Part of the Answer: Examining the Impact of Emotion Self-reports on Student Emotion. In *User Modeling, Adaptation, and Personalization* (pp. 471-477). Springer International Publishing.

Wixon, M., Baker, R.S.J.d., Gobert, J., Ocumpaugh, J., Bachmann, M. (2012) WTF? Detecting Students who are Conducting Inquiry Without Thinking Fastidiously. *Proceedings of the*

*20th International Conference on User Modeling, Adaptation and Personalization (UMAP 2012)*, 286-298.

Wolfgang, A., & Cohen, M. (1988). Sensitivity of Canadians, Latin Americans, Ethiopians, and Israelis to interracial facial expressions of emotions. *International Journal of Intercultural Relations*, *12*(2), 139-151.

Wragg, T. (2013). *An introduction to classroom observation*. Routledge.

Yohalem, N. & Wilson-Ahlstrom, A, with Fischer, S., & Shinn, M. (2009, January). Measuring Youth Program Quality: A Guide to Assessment Tools, Second Edition. Washington, D.C.: The Forum for Youth Investment, Impact Strategies.

Zeman, J., Klimes-Dougan, B., Cassano, M., & Adrian, M. (2007). Measurement issues in emotion research with children and adolescents. *Clinical Psychology: Science and Practice*, *14*(4), 377-401.

Zuckerman, M., Eysenck, S. B., & Eysenck, H. J. (1978). Sensation seeking in England and America: cross-cultural, age, and sex comparisons. Journal of consulting and clinical psychology, 46(1), 139.