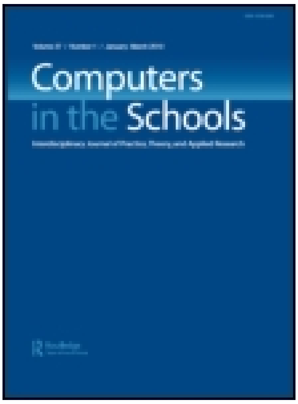

Computers in the Schools
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wcis20

Evaluation of Handheld Computers for Direct Systematic Classroom Observation

Michael W. Bahr [a], Donald A. Gouwens [a] & Genevieve Schuh [b]
[a] University of Missouri-St. Louis, St Louis, Missouri, USA
[b] St. Louis Special School District, St Louis, Missouri, USA
Published online: 13 Sep 2012.

Computers in the Schools, 29:268–284, 2012

ISSN: 0738-0569 print / 1528-7033 online
DOI: 10.1080/07380569.2012.702720

Routledge
Taylor & Francis Group

Evaluation of Handheld Computers for Direct Systematic Classroom Observation

MICHAEL W. BAHR and DONALD A. GOUWENS
University of Missouri–St. Louis, St Louis, Missouri, USA
GENEVIEVE SCHUH
St. Louis Special School District, St Louis, Missouri, USA

Through this study the authors evaluate outcomes associated with the use of handheld computers by interventionists in improving the efficiency of direct systematic classroom observation. Information from observations is used by interventionists for treatment planning and evaluation. In this study, interventionists were trained to use personal digital assistants with classroom observational software for use with students who displayed low levels of academic engagement. Results indicated that the personal digital assistants and observational software were perceived as user-friendly, increased computer self-efficacy, and facilitated treatment planning and evaluation. Discussion focuses on implications for use of handheld computers and mobile devices by interventionists.

KEYWORDS *assessment, classroom observation, technology, intervention planning*

With the increasing emphasis on data-based decision-making in education (Brown-Chidsey & Steege, 2010), coupled with advances to portable technologies in data collection and data management, school-based interventionists—teacher consultants, counselors, school psychologists—are turning to handheld computers and mobile devices to assist with the planning and evaluation of student interventions. Whereas the use of handhelds

This research was supported by an Innovation Grant awarded to the first author by the University of Missouri-St. Louis.
The authors express their gratitude to Joseph Bryant for his assistance with data analysis.
Address correspondence to Michael W. Bahr, PhD, Division of Education Psychology, Research and Evaluation, One University Blvd., University of Missouri-St. Louis, St. Louis, MO 63121. E-mail: bahrmi@umsl.edu

by teachers has occurred for over a decade (Warschauer, 2011), the adoption of portable technologies by other school professionals has followed at a somewhat slower pace (Adiguzel, Vannest, & Zellner, 2009).

Slower adoption is puzzling given the many advantages of handheld computers and mobile devices for interventionists. Foremost, whereas classroom-based assessments, such as observation of student behavior, have traditionally been conducted using paper-and-pencil, handhelds provide an efficient, less cumbersome method for recording data (Bennett & Cunningham, 2009). The smaller size and increased portability offer significant advantages compared to even light-weight laptops (Adiguzel, Vannest, & Parker, 2009), all with storage capacity that far exceeds the needs of daily data collection (Fletcher, Erickson, Toomey, & Wagenaar, 2003). Additionally, research (Olswag, Svensson, Coggins, Beilinson, & Donaldson, 2006) has demonstrated that observers can be trained to observe accurately and achieve adequate to high levels of reliability, thereby improving the overall precision of data collection.

DIRECT SYSTEMATIC CLASSROOM OBSERVATION

Given these advantages, handhelds appear ideally suited to assist with data collection. Data-based decision-making by interventionists provides an objective assessment methodology to develop, implement, and evaluate school interventions, thus reducing reliance on subjective judgments by teachers and parents about student progress. Direct, systematic observation of student behavior represents one of the most frequently used data-based techniques for students with academic and behavior problems (Hintze, Volpe, & Shapiro, 2008; Wilson & Reschly, 1996), and it represents an area where technological applications are beginning to assist in making data-based decisions (Silberglitt, 2008).

Observation is considered direct when a behavior is defined in clear, measureable terms prior to an observation and thus is capable of being readily monitored and quantified. Observation is systematic when it is conducted in a standardized fashion using research-based techniques, such as interval recording or momentary time sampling (Hintze et al., 2008).

An example of direct systematic observation is the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2011), which is designed to assess student engagement in the classroom. Academic engagement is defined as the amount of time a student is actively engaged in learning tasks. Research over the past three decades has demonstrated a moderate to high correlation between academic engagement and student achievement (Gettinger & Ball, 2008). "This link between time and learning is one of the most enduring and consistent findings in educational research" (Gettinger & Ball, 2008, p. 1043). The primary purpose of the BOSS is to allow interventionists to observe and record student engagement using a direct systematic observation protocol. Information from BOSS observations contributes to intervention planning and evaluation. Initially developed as a paper-and-pencil recording system, BOSS software was created for the PalmPilot Personal Digital Assistant (PDA); and development of updated applications for mobile devices, including the iPad, iPhone, and Android, is underway (Shapiro, 2011).

Psychometrically, the BOSS has demonstrated satisfactory reliability and validity. For example, the BOSS possesses high interobserver reliability with Kappa coefficients consistently above .90 (Volpe, DiPerna, Hintze, & Shapiro, 2005). The BOSS has demonstrable discriminate validity in successfully differentiating clinical groups (e.g., children with attention-deficit/hyperactivity disorder) from normal groups (DuPaul et al., 2004), and treatment validity was apparent by increasing academic engagement through instructional interventions (Ota & DuPaul, 2002). In a review of seven observation systems, Volpe and colleagues (2005) favorably noted how the BOSS incorporates observation of both positive behavior (i.e., academic engagement) and problem behaviors (i.e., off-task).

Observation Categories

The BOSS contains three observational categories, two of which focus on student behavior and one on teacher behavior. The first student category is academic engagement, which has two subcategories, active engagement and passive engagement. *Active engagement* involves a student verbally responding to the teacher, talking in a small instructional group, or writing (e.g., taking notes). *Passive engagement* includes listening to the teacher, looking at instructional materials (e.g., reading a book), or listening to other students recognized by the teacher.

The second observational category is off-task and contains three subcategories. *Verbal off-task* is inappropriate talking (e.g., talking to a peer, talk outs). *Off-task motor* involves any example of motor activity not associated with an instructional task, such as walking around the room, playing with an object (e.g., pencil or paper clip), or drawing or writing that is not related to an assignment. *Off-task passive* is defined as three consecutive seconds when a student is not looking at the teacher, instructional materials, or another student who is responding to an instructional question or discussion. This includes behaviors such as daydreaming, staring out the window, or looking around the room.

The third observational category is *teacher-directed instruction*, defined as the teacher providing instruction to the class, a group, or an individual. It includes teaching a lesson, working at the board, or instructing an individual student; it excludes grading papers, preparing instruction, or speaking to someone about a non-instructional issue.

In addition to the observation categories, student information (e.g., name, grade) is entered as well as general information about the observation (e.g., date, class activity/subject, larger vs. small group instruction).

Observation Protocol

BOSS classroom observations are typically 15–30 minutes in length, and each minute is divided into 15-second observation intervals. Two different observation techniques—momentary time sampling and partial interval recording—are used for systematic observation. Momentary time sampling is used to observe and record academic engagement. This technique requires the observer to look at the target student at the beginning of each 15-second interval and codes whether the student is academically engaged. If so, the observer codes either active engagement or passive engagement, whichever is appropriate. For the remainder of the 15-second interval, the observer uses partial interval recording to record any instance of off-task behavior—verbal, motor, or passive. More than one type of off-task behavior may be recorded, and because partial interval recording is used, the off-task behavior need only occur during part of the 15-second interval, not the entire interval.

This observation process is repeated for each 15-second interval with one exception. At every fifth interval, the academic engagement and off-task behavior of a peer student are observed using momentary time sampling and partial interval recording, respectively. This provides a set of comparison data between the target student and a peer. In addition, using partial interval recording, teacher-directed behavior is observed and recorded every fifth interval as well.

To illustrate the use of the BOSS in treatment planning and evaluation, Table 1 displays the results of two 20-minute observations using the PDA/BOSS to observe Caleb, a fourth-grader, who was referred for attention problems and poor work in math. Caleb's teacher was particularly concerned about Caleb's inattention when written seatwork was assigned. The first (or preintervention) observation summary shows that Caleb was academically engaged for 34.4% (26.6% active + 7.8% passive) of the observation intervals. By contrast, a peer was engaged 87.5% (62.5% active + 25% passive) of the observation. The concern about Caleb's discrepant level of engagement, which was 53.1% lower than the peer, is exacerbated by his high levels of off-task motor behavior (39.1% vs. 6.3% for peer) and off-task passive behavior (21.9% vs. 0% for peer). This information was used to develop a self-monitoring intervention designed to increase Caleb's attention and work productivity during independent seatwork. The postintervention observation completed approximately 8 weeks later demonstrated two important points. First, Caleb's relatively low level or academic engagement was a primary consideration in developing a treatment plan to improve attention and work

TABLE 1 Caleb's BOSS Pre- and Postintervention Summaries with Identifying Information and Frequencies and Percentages of Observed Behavior

Preintervention Observation

Student: Caleb	Observer: Jessica P.	Grade: 4
Date: 1/17/2011	Time: 10:42:08 AM	Task: Math
Setting: ISW:TPsnt	Duration: 20 Minutes	Interval: 15 Seconds

	Target student		Peer student		Teacher	
Behavior	Count	Percentage	Count	Percentage	Count	Percentage
AET	17	26.6%	10	62.5%	—	—
PET	5	7.8%	4	25.0%	—	—
OFT-M	25	39.1%	1	6.3%	—	—
OFT-V	0	0.0%	1	6.3%	—	—
OFT-P	14	21.9%	0	0.0%	—	—
TDI	—	—	—	—	1	6.3%
	Target intervals: 64		Peer/teacher intervals: 16			

Postintervention Observation

Student: Caleb	Observer: Jessica P.	Grade: 4
Date: 3/22/2011	Time: 10:48:17 AM	Task: Math
Setting: ISW:TPsnt	Duration: 20 Minutes	Interval: 15 Seconds

	Target student		Peer student		Teacher	
Behavior	Count	Percentage	Count	Percentage	Count	Percentage
AET	46	71.9%	11	68.7%	—	—
PET	9	14.1%	3	18.8%	—	—
OFT-M	2	3.1%	2	12.5%	—	—
OFT-V	0	0.0%	1	6.3%	—	—
OFT-P	7	10.9%	0	0.0%	—	—
TDI	—	—	—	—	3	18.8%
	Target intervals: 64		Peer/teacher intervals: 16			

Note. AET = Actively engaged in task; PET = Passively engaged in task; OFT-M = Off-task motor; OFT-V = Off-task verbal; OFT-P = Off-task passive; TDI = Teacher-directed instruction; Setting ISW:TPsnt = Student in independent seatwork, teacher present.

productivity. Second, the BOSS data were useful in evaluating intervention effectiveness. Postintervention results indicated Caleb's engagement of 86% (71.9% active + 14.1% passive) was comparable to a peer's level of 87.5% (68.7% active + 18.8% passive), and his off-task behaviors decreased as well.

When coupled with other sources of information, observation data are useful to interventionists and teachers who design instructional modifications for struggling students. In contrast to the laborious hand-scoring required for the paper/pencil version of the BOSS, the PDA software immediately and accurately calculates summary statistics for an observation.

CURRENT STUDY

In the current study we evaluated the overall utility of PDAs in conducting direct, systematic observation. Although previous research (see Volpe et al., 2005) on the BOSS indicated satisfactory psychometric adequacy, no research exists on the utility of PDAs with the BOSS software. More importantly, we attempted to address the dearth of research associated with the use of handhelds by interventionists.

METHOD

Participants

Applied specialties in psychology (e.g., clinical or school) are small, graduate-intensive, training programs, and the participants were 25 graduate candidates, 22 of whom were enrolled in an Educational Specialist in School Psychology degree program in the Midwest. One candidate was enrolled in school counseling, and two were taking classes for professional development. Participants were predominantly female (84%) and Caucasian (96%). Of the 22 enrolled in the program, half were employed as school-based practitioners (9 psychological examiners, 1 social worker, 1 teacher).

Eleven (44%) were younger than 30 years of age with the remaining 14 participants age 30 or older, including 3 who were older than 50 years of age. Eleven (44%) held a bachelor's degree with 10, 2, and 1, respectively, holding master's, educational specialist, or doctoral degrees. The range of professional experience in school varied considerably ($M = 4.45$, $SD = 6.78$, $Mdn = 2.00$) with 5 participants with 0 years' experience to 1 with 30 years. Seven (29%) participants had previously used the paper/pencil version of the BOSS, and none had used the PDA version. Instruction on direct systematic observation took place in one of two courses. Eleven participants were enrolled in a course titled Psychotherapeutic Interventions, which focused on treatments for social, emotional, and behavior problems. Fourteen participants were enrolled in a course titled Psychoeducational Interventions, which focused on assessment and intervention of academic skill problems. The 25 participants took these classes at separate times, and no participant was in both classes at the time of the study.

Measures and Instruments

Three outcome measures were used to evaluate outcomes on use of the PDA and BOSS: data on PDA usage, self-efficacy, and user satisfaction.

Usage

The first variable was a usage index that simply served as a "count" for how many times a participant used the PDA for a BOSS observation. Participants had some flexibility in using the PDA. A minimum of two BOSS observations were required for a course case study (described hereafter), but participants could use the PDA more if they wished. The counts were taken by checking each PDA and counting the number of observations by a participant.

Self-efficacy

The Computer Self-Efficacy Scale (Brinkerhoff, 2006), a 20-item self-report measure, was used to assess self-efficacy. This measure was initially developed by Cassidy and Eachus (2002) and modified by Brinkerhoff for a series of program evaluations with educators. Cassidy and Eachus (2002) reported acceptable reliability and construct validity on this measure. Brinkerhoff (2006) reduced the number of items from 30 to 20, maintained a balance of 10 positively keyed items and 10 negatively keyed items, and demonstrated adequate reliability (Cronbach alpha = .94). Some of the items on this measure included "I am very confident in my abilities to use computers. Computers make me much more productive; I usually find it easy to learn how to use a new software package. I consider myself a skilled computer user." Participants in the current evaluation used the 20-item version and rated each item on a 4-point scale: 4 (*strongly agree*), 1 (*strongly disagree*).

User Satisfaction Survey

To evaluate user satisfaction, a survey was created with three sections. The first section contained 10 Likert-type items that asked about the PDA or the BOSS. Specifically, questions included how helpful the PDA was, how easy or difficult it was to learn either the PDA or BOSS, and a series of questions on the extent to which the BOSS facilitated activities such as data collection or treatment planning. The second section contained three open-ended questions/statements: (a) What were some advantages of using the PDA and BOSS? (b) What were the limitations associated with using the

PDA and BOSS? (c) Please provide any other comments about the utility of using the PDA/BOSS. These were written responses. The third section was a demographic sheet.

Instruments

Two instruments were necessary for conducting this evaluation. The first instrument was the PDA. Twelve Palm Z22 PDAs were purchased at the cost of $99.00 per unit. This Z22 model was considered a basic PDA with 32MB of memory, and it was capable of running the BOSS software. The second was the purchase of eight BOSS software programs. This software was $79.00 per unit from Harcourt Assessment, Inc. A unit included a CD to load two BOSS programs onto separate computers.

Design

This study was designed as a program evaluation of PDAs for direct systematic observation. Given that participants were in different graduate classes—albeit, both with the same instructor using the same instructional procedures on the PDA/BOSS—it seemed appropriate to examine possible differences by course. Thus, in addition to descriptive, inferential, and qualitative analyses on group outcomes, between-group analyses by course subsamples were also completed. The term *preintervention* refers to the beginning of the semester prior to instruction on the use of PDAs and the BOSS. *Postintervention* pertains to the end of the semester after instruction on and use of the PDA/BOSS.

Procedures

Preintervention Data Collection

At the beginning of each semester, candidates were invited to participate voluntarily in the study and informed that course grades were not linked to participation. After consenting to participate, the self-efficacy scale was completed.

Instruction on Classroom Observation/PDA

Each of the classes from which participants were recruited met once a week for 3 hours over 15 weeks. During the second and third class sessions for each course, general instruction on best practices in classroom observation was presented, followed by specific instruction in using PDAs and the BOSS. Prior to each instructional session, participants were required to complete assigned readings on direct systematic observation and from the BOSS manual (Shapiro, 2011).

Instruction specifically comprised two components. The first component focused on general issues about the importance of classroom observation in assessment and treatment, explanation of basic principles (e.g., linking assessment to intervention) and concepts (e.g., target behavior), and instruction on observational strategies (e.g., interval recording) and when and why to use them.

The second component focused on use of PDAs and the BOSS. Instruction on the BOSS addressed whom to observe (i.e., target vs. peers vs. teacher), definition of categories of observed behavior (i.e., academic engagement, off-task behavior, and teacher-directed instruction), and the observation methods (i.e., momentary time-sampling, partial interval recording). This component also included a "how to" with the PDA and provided instruction on basic steps (e.g., turning on the PDA, selecting the BOSS software from the menu) and advanced activities (i.e., creating an observation file by entering student demographic information). In-class practice allowed candidates to develop proficiency in using the BOSS/PDA and downloading observational data into Word files for student reports.

Case Study

The PDA was used as part of a course project—the case study—and this involved a participant working as a school consultant to a teacher who was concerned about an elementary school student with significant academic deficits or a severe behavior problem. Participants worked with teachers over an 8-week period using the problem-solving consultation process (Kratochwill, 2008) to improve the academic performance or classroom behavior of the student by developing an intervention plan and evaluating it. The case study culminated in a written report summarizing the consultation, and it contained a graph with data on the student's academic or behavioral progress in response to a classroom intervention developed by the consultant and teacher. The PDA/BOSS was used primarily in two parts of the problem-solving process: the problem identification stage, where it was used to collect data on the student's behavior, and the treatment evaluation stage, where it was used as an evaluative measure.

Post Data Collection

At the conclusion of the semester, PDA usage data were collected, and participants were invited to complete the self-efficacy scale again and the user satisfaction survey.

Data Analysis

The Computer Self-Efficacy Scale had 10 items that were rekeyed so that all 20 items would be stated affirmatively, with higher scores indicating greater self-efficacy. Next, as a check on the psychometric adequacy of this scale, reliability analyses were conducted using the preintervention scores ($N = 25$), and a Cronbach alpha of .95, comparable to previous research (Brinkerhoff, 2006), was obtained. The total score on this measure was then calculated.The usage data and the self-efficacy scores were analyzed using inferential statistics.

The rated items on the user satisfaction survey were analyzed descriptively; that is, by inspecting their means and standard deviations and noting the relatively higher versus lower ratings. Statistical analyses were not used because these items were not associated with a psychometrically validated instrument.

The three open-ended items were analyzed in two ways. First, responses were qualitatively evaluated and placed in categories described hereafter. Second, once the categories were established, nonparametric analyses were used, where appropriate, to evaluate differences in the frequency with which statements were identified by category.

To develop categories, two members of the research team began by reading all written responses to each question. These were reread several times and discussed extensively until various themes emerged. Eventually, categories into which a written response could be placed were developed. A written response was defined as a statement that contained a meaningful piece of information. For example, the sentence, "The PDA was really easy to use" was one statement. The sentence, "It took time to learn to use the BOSS, but its automatic scoring made it worthwhile" was counted as two statements (i.e., took time to learn, automatic scoring is a plus).

For the first open-ended question on advantages of the BOSS/PDA, two categories were established: features and ease of use. *Features* was defined as a statement that focused on specific functions of the BOSS software such as automatic scoring, generation of statistics or student profiles on level of academic engagement, and date storage. *Ease of use* was defined as any statement pertaining to how easy the BOSS/PDA was to use, such as efficient, quick, or helpful.

For the second question on disadvantages, two categories were established: volume/signaling and other problems. *Volume/signaling* was defined as a problem that concerned the volume of the PDA and the BOSS signal to notify the observer to observe and record. *Other problems* was defined as an issue not related to volume/signaling.

The last question invited any other comments about the utility of the BOSS/PDA, and two categories emerged: ease of use and other comments. The definition for *ease of use* was the same as the previous definition for this category in the first question. Any statement that did not pertain to ease of use was categorized into *other comments*.

When the two researchers were confident that the categories were clearly defined and were able to classify items consistently and accurately, a rater who was blind to the study was trained by one of the researchers to assign items to categories. This was done by teaching the blind rater the categories associated with each item and then having the blind rater classify practice items, which were similar responses to the actual data. When the blind rater reached a high level of proficiency with the practice items, 50% of the actual data was given to the blind rater to classify according to the established categories. The blind rater's classification of responses by category was compared to those of one of the researchers, who served as the criterion rater. Interobserver agreement was then determined by calculating percentage agreement (number of correct classifications by blind rater divided by the correct plus incorrect classifications and multiplied by 100). Across all categories, the range of agreement was 83%–100%, with a median of 100%.

RESULTS

Outcomes for Total Group

Usage

The average usage of the PDA was 2.95 ($SD = 1.57$), indicating that participants used the BOSS more than the required two-observation minimum.

Self-Efficacy

Total scores on the self-efficacy measure could range from a low of 20 to a high of 80. A paired t test revealed that the 25 participants significantly increased self-efficacy from preintervention ($M = 62.40$, $SD = 9.83$) to postintervention ($M = 65.92$, $SD = 10.08$), $t\ (24) = 3.99$, $p = .001$, Cohen's $d = .35$).

User Satisfaction Survey

The user satisfaction survey contained (a) ratings on various aspects of the PDA or the BOSS/PDA, and (b) responses to open-ended questions.

Ratings

Table 2 provides descriptive statistics for the opinion ratings from the user satisfaction survey. Because these items were not associated with a psychometrically validated measure, their analysis and interpretation were completed descriptively. At least three trends were apparent in the satisfaction ratings. First, in general, participants assigned positive to very positive ratings for all items ($Mdn = 4.39$). Second, items 1–3, focusing on the PDA

TABLE 2 Descriptive Statistics on User Satisfaction Survey (N = 24)

Items	*M* (*SD*)
Overall, how helpful was the PDA? (5 = *Very helpful*, 1 = *Not very helpful*)	4.52 (0.99)
How easy or difficult was it to learn how to use the PDA? (5 = *Very easy,* 1 = *Very difficult*)	4.45 (0.93)
How likely is it that you'll recommend using a PDA to other psychologists/counselors? (5 = *Very likely,* 1 = *Not very likely*)	4.41 (1.05)
How easy or difficult was it to learn how to use the BOSS on the PDA? (5 = *Very easy,* 1 = *Very difficult*)	4.62 (0.49)
Overall, to what extent did use of the BOSS/PDA contribute to your professional development as a psychologist or counselor? (5 = *Quite a bit,* 1 = *Not very much*)	3.95 (1.02)
I think the BOSS/PDA facilitated data collection. (5 = *Strongly agree,* 1 = *Strongly disagree*)	4.60 (0.72)
The BOSS/PDA facilitated treatment planning. (5 = *Strongly agree,* 1 = *Strongly disagree*)	3.95 (1.14)
The BOSS/PDA facilitated my report writing. (5 = *Strongly agree,* 1 = *Strongly disagree*)	4.17 (1.11)
The BOSS/PDA will make my work with children and adolescents more efficient. (5 = *Strongly agree,* 1 = *Strongly disagree*)	4.37 (0.87)
Overall, the BOSS/PDA will improve the quality of my work with children and adolescents. (5 = *Strongly agree,* 1 = *Strongly disagree*)	4.20 (0.88)

Note: All ratings were conducted using a Likert-type, 5-point rating scale with higher numbers indicating more positive responses

only, revealed high, positive ratings (*Mdn* = 4.45), indicating that participants found the PDA to be very helpful, easy to learn, and likely to be recommended to colleagues. Third, items 4–10, which addressed ease and utility of the PDA/BOSS, were in a very positive range as well (*Mdn* = 4.20), indicating that it was easy to learn (*M* = 4.62, highest rated item overall) and facilitated important activities such as data collection, treatment planning, and report writing.

Open-Ended Questions

The first analysis on the three open-ended questions examined whether differences existed on the frequency of responses by question. The total number of written comments for advantages, disadvantages, and general comments on PDA/BOSS utility were 54, 30, and 18, respectively. Chi-square analysis, χ^2 (2) = 19.76, $p < .001$, was significant, and the distribution of scores indicated there were significantly more responses to the advantages question compared to the disadvantages and utility questions.

Of the 54 comments on advantages, 35 statements were in the features category, which was significantly more than the remaining 19 in the ease of use category, χ^2 (1) = 4.74, $p < .05$. The PDA/BOSS features mentioned most often were the ability to do target student versus peer comparisons, notification by the PDA/BOSS for when to observe, automatic scoring, summarization of data in tables, and storage of data. Of the 19 statements in the ease of use category, 8 actually used the phrase "easy to use" or a close derivative. Other advantages included the PDA/BOSS's ability to promote more efficient work by professionals in classrooms, its size (small) and lack of paper (only one item needed to do an observation), its portability (I could walk around the room to observe), and how it allowed an observer to be inconspicuous. One respondent wrote, "[the] teaching staff was envious of the ease" with which data could be collected on students.

Of the 30 comments on disadvantages, 21 were in the other problems category, and this was significantly more than the 9 in the volume/signaling category, χ^2 (1) = 4.80, $p < .005$. The most frequent issue identified in other problems was a concern about the inability to take clinical notes on important child behaviors that occurred during a classroom observation but were not included in the BOSS's observational system. This was followed by comments saying the BOSS was less applicable for academic skill deficits (than for behavior problems). Comments in volume/signaling all identified a problem with the audio signal and vibrate functions of the PDA/BOSS. The BOSS software comes equipped with both auditory and visual signals to cue an observer to observe and then record data. Most PDAs have a vibrate function as well. However, the auditory/vibrate functions were not a feature of the Palm Z22, so there was only a visual cue to begin an observation interval. This was a hardware, not a software (i.e., BOSS), problem.

Eighteen comments were made in response to the third question—other comments on the utility of the PDA/BOSS. Ten and 8, respectively, were categorized into ease of use and other comments, and there was no difference in the frequency of responses into these categories, χ^2 (1) = .22, *ns*. Responses to this question reiterated how easy it was to use the PDA/BOSS. Comments included, "I have taught two other people at my site how to use it," "I'd like to have [a PDA/BOSS]—it's infinitely superior to other methods," and "I have already recommended [the PDA/BOSS to my director], and my practicum supervisor is even interested in it." A response in the other comments category included one more endorsement: "I have sat in meetings and shared my BOSS findings with other staff, and they are always impressed."

Usage Outcomes by Course

One additional set of analyses on usage were conducted by subsamples based on age, highest degree obtained, years of experience, practitioner status (yes vs. no), and course. Results indicated differences on the course subsample only. Since the PDA/BOSS instructional component was implemented in different treatment courses—Psychotherapeutic Interventions, Psychoeducational Interventions—it was important to understand how outcomes varied. Results on usage indicated that 10 participants from the Psychotherapeutic Interventions course had a total of 43 observations (M = 4.30, SD = 1.70), and 14 participants in the Psychoeducational Interventions course had a total of 28 observations (M = 2.00, SD = 0.00). An independent t test showed a significant difference, t (9) = 4.27, p = .01 (*df* were adjusted because variances were unequal).

DISCUSSION

This study investigated the utility of handheld computers with observational software and their impact on performance outcomes associated with treatment planning and treatment evaluation. It also examined performance outcomes by the type of course in which participants were enrolled.

In terms of the broad utility of the PDA/BOSS, multiple sources of outcome data suggest that the PDA/BOSS was useful to interventionists. For example, participants' self-evaluation of computer efficacy improved from pre to post when using the PDA/BOSS. User satisfaction responses indicated that participants perceived the PDA as a helpful clinical instrument that was easy to learn to use. Similarly, user satisfaction ratings indicated that the combination of the PDA and BOSS was easy to learn to use, facilitated data collection and treatment planning, and increased the efficiency with which participants worked with children and adolescents.

Written responses on the user satisfaction survey reinforced participant conclusions about how easy the PDA/BOSS was to use. More specifically, participants noted how the BOSS improved the efficiency of classroom observations by using objective categories for observation of student behavior and by using a small PDA instead of a notepad, pencil, and stopwatch. Participants commented on how quickly observations could begin and end by using a PDA. Closely related to ease of use was the participants' positive evaluation of the features of the PDA/BOSS. These included the preset intervals for observing and recording behavior, inclusion of peer observations for comparison purposes, the automatic calculation of the percentage of displayed behavior by category when an observation ended, and the ability to download data summaries into a student report. These strengths are consistent with those identified in Fletcher's research (Fletcher et al., 2003) comparing PDA versus paper-and-pencil data collection methods.

The two notable limitations of the PDA/BOSS identified by participants included the lack of an audible signal (or vibration) to begin an observation interval and the inability to write additional notes about classroom behavior during the observation. A recommendation to address the former concern is to ensure the handheld device includes a clearly audible signal or vibrate function; the latter concern is simply remedied by observers keeping a notepad available during observations. These concerns notwithstanding, responses to user satisfaction items, both ratings and open-ended questions, were overwhelmingly positive about the PDA/BOSS and, coupled with the improved computer self-efficacy, provide additional evidence of the utility of handheld computers and mobile devices in clinical work with children and adolescents.

The subsample analysis by course revealed a higher usage pattern by participants in the Psychotherapeutic Interventions class. A plausible explanation for higher usage is that disengagement (i.e., inattention) is a corollary of social-emotional or behavior problems. Consequently, the PDA/BOSS can be very useful for problem identification and problem analysis, and in some of the case studies, it was used in treatment evaluation as a dependent variable to assess changes in student behavior due to an intervention. By contrast, participants in the Psychoeducational Interventions course worked on cases where academic skill deficits were the focus of the intervention (e.g., reading performance), and progress monitoring of outcomes typically occurred using a curriculum-based measure such as reading fluency. In these cases, the PDA/BOSS may be used to rule out disengagement or inattention as the main problem, thereby serving as a secondary data source to verify an academic deficit as the primary concern.

Success with the paper/pencil version of the BOSS relies on substantial practice because the observer must monitor interval length (usually with a stopwatch), observe at the appropriate moment, and then record engagement and off-task behaviors (Shapiro, 2011). By contrast, handheld technologies automatically regulate interval length, queue the observer who to observe and when, and allow recording via a simple touch of the screen. Ease of use has been identified as an important criterion for technology adoption by professionals (Davis, Bagozzi, & Washaw, 1989), and the results of this program evaluation provide additional confirmation of how handheld computers and mobile devices are changing a well-developed but somewhat cumbersome observation protocol into an efficient and user-friendly endeavor.

These findings should be considered relative to study limitations. These include the use of two graduate classes that provided intact groups but prohibited randomized assignment. Although both groups were exposed to the same instruction on the PDA/BOSS, one group used it for clinical work for students with behavior problems; whereas the other group used it with students possessing academic deficits. Significant pre/post differences were observed for usage and computer self-efficacy, but the study's design (i.e., program evaluation vs. experimental design) precluded comparisons with a control group, which cannot rule out changes due to maturation or other factors. Lastly, as noted, some of the outcome data were analyzed descriptively, which necessitates cautious interpretation of those data.

CONCLUSION

According to the National Education Technology Plan (U.S. Department of Education, 2010), effective learning in the 21st century will be contingent upon new and varied assessment methodologies to identify student strengths and deficits, and "technology-based assessments can provide data to drive decisions on the basis of what is best for each and every student" (p. vii). The utility of handheld technologies for direct systematic observation appears to be consistent with the emphasis on data-based decision-making espoused by the Department of Education. The results of this evaluation provide important evidence for the utility of software for handheld computers and mobile device applications for interventionists and their work with children and adolescents. The major conclusions indicated that the handheld computers are easy to use, possess features that increase the efficiency and quality of clinical work, enhance computer self-efficacy, and are perceived positively by practicing interventionists and clinicians in training.

REFERENCES

Adiguzel, T., Vannest, K. J., & Parker, R. I. (2009). Are handheld computers dependable? A new data collection system for classroom-based observations. *Journal of Special Education Technology*, *24*, 31–46.

Adiguzel, T., Vannest, K. J., & Zellner, R. D. (2009). The use and efficacy of handheld computers for school-based data collection: A literature review. *Computers in the Schools*, *26*, 187–206.

Bennett, K. R., & Cunningham, A. C. (2009). Teaching formative assessment strategies to preservice teachers: Exploring the use of handheld computing to facilitate the action research project. *Journal of Computing in Teacher Education*, *25*, 99–105.

Brinkerhoff, J. (2006). Effects of a long-duration, professional development academy on technology skills, computer self-efficacy, and technology integration beliefs and practices. *Journal of Research on Technology in Education*, *39*, 22–43.

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice* (2nd ed). New York, NY: Guilford.

Cassidy, S., & Eachus, P. (2002). Developing the Computer User Self-Efficacy (CUSE) Scale: Investigating the relationship between computer self-efficacy, gender, and experience with computers. *Journal of Educational Computing Research*, *26*, 133–153.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*, 982–1003.

DuPaul, G. J., Volpe, R. J., Jitendra, A. K., Lutz, J. G., Lorah, K. S., & Grubner, R. (2004). Elementary school students with attention-deficit/hyperactivity disorder: Predictors of academic achievement. *Journal of School Psychology*, *42*, 285–301.

Fletcher, L. A., Erickson, D. J., Toomey, T. L., & Wagenaar, A. C. (2003). Handheld computers: A feasible alternative to paper forms for field data collection. *Evaluation Review*, *27*, 165–178.
Gettinger, M., & Ball, C. (2008). Best practices in increasing academic engaged time. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1043–1057). Bethesda, MD: National Association of School Psychologists.
Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2008). Best practices in the systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 319–335). Bethesda, MD: National Association of School Psychologists.
Kratochwill, T. R. (2008). Best practices in school-based problem-solving consultation: Applications in prevention and intervention systems. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1673–1688). Bethesda, MD: National Association of School Psychologists.
Olswag, L. B., Svensson, L., Coggins, T. E., Beilinson, J. S., & Donaldson, A. L. (2006). Reliability issues and solutions for coding social communication performance in the classroom. *Journal of Speech, Language, and Hearing Research*, *49*, 1058–1071.
Ota, K. R., & DuPaul, G. J. (2002). Task engagement and mathematics performance in children with attention-deficit hyperactivity disorder: Effects of supplemental computer instruction. *School Psychology Quarterly*, *17*, 242–257.
Shapiro, E. S. (2011). *Academic skills problems workbook* (rev. ed.). New York, NY: Guilford.
Silberglitt, B. (2008). Best practices in using technology for data-based decision making. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 1869–1883). Bethesda, MD: National Association of School Psychologists.
U.S. Department of Education. (2010). *Transforming American education: Learning powered by technology*, *National Education Technology Plan*. Washington, DC: Department of Education.
Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, *34*, 454–474.
Warschauer, M. (2011). *Learning in the cloud: How (and why) to transform schools with digital media*. New York, NY: Teachers College Press.
Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology. *School Psychology Review*, *25*, 9–23.