DEVELOPMENT AND RELIABILITY OF AN OBSERVATION SYSTEM: PERSONAL DIGITAL ASSISTANT (PDA) SYSTEMATIC OBSERVATION SOFTWARE (SOS): A PILOT STUDY

Thaynan L. Knowlton
Kristofer H. Cruz
Eastern Washington University

ABSTRACT

As the role of the school psychologist expands, practitioners will need to utilize alternative methods of assessment. Systematic observation is one example of an important tool that could be applied in a variety of settings. For this pilot study, a new observation instrument was developed and tested for reliability. To accomplish this, twenty-five minutes of videotaped classroom activity from the Ecobehavioral Assessment Systems Software (EBASS) calibration regular education videotape was shown to participants. Participants recorded verbal/motor/passive on-task and off-task behaviors for a student in a fifth grade regular education classroom. Participants (N=17) consisted of school psychologists, special education personnel and graduate level psychology students. It is demonstrated that an interval recording version of Systematic Observation Software (SOS) designed for the Newton has acceptable reliability. Results, using kappa, indicate an overall interobserver reliability coefficient of .73.

DEVELOPMENT AND RELIABILITY OF AN OBSERVATION SYSTEM: PERSONAL DIGITAL ASSISTANT (PDA) SYSTEMATIC OBSERVATION SOFTWARE (SOS): A PILOT STUDY

The four pillars of assessment as identified by Sattler (1992) include norm-referenced tests, interviews, informal assessment, and observations. Systematic observations fall into one of four commonly used methods: event, narrative, rating scale, and interval. Each type of observation is used in a specific circumstance, and each offers its own set of unique advantages and disadvantages. The purpose of this paper is to describe the development and reliability of a new, systematic observation tool.

According to a National Association of School Psychologists (NASP) survey done in 1994, school psychologists' time was reportedly spent in the following ways: assessment related activities 51%, consultation 20%, treatment 19%, and other activities (i.e., administrative duties 6%, continuing education 3%, and research 1%) (Stinnett, Havey & Oehler-Stinnett, 1994). During the last 20 years there has been increasing criticism of standardized testing, and many are calling for a more expanded role for the school psychologist (Batsche, 1992; Naglieri, Das, & Jarman, 1990; Graden, Zins, & Curtis, 1988; and Reschly, 1980, 1988). However, because of current federal law, funding structures have firmly put the psychologist in the role of assessing students for placement.

Stinnett, Havey & Oehler-Stinnett's NASP survey (1994) reported that in the behavior-social-emotional assessment domain, 54% of surveyed school psychologists use direct behavior observation in their settings. It was also given an importance rating of 2.80 (3 = very important, 2 = somewhat important, and 1 = not important) which was among the top 10% of common assessment tools overall (i.e., WISC-III: 2.91, WISC-R: 2.89, WAIS-R: 2.82, teacher interview: 2.82, and child interview: 2.81). These results are contrasted with such standardized observation protocols as the CBCL (35%; 2.55), Conners (38%; 2.47), and other widely recognized observation measures.

In terms of frequency (number of administrations completed by the respondents in a typical school year), behavior observation scored second highest of all measures in the 1994 sample (behavior observation: M=59.93 vs. teacher interview: M=60.31). These figures are also contrasted with the highest frequencies from tests in six of the domains surveyed: Intellectual-Cognitive (WISC-R: 45.58), Academic Achievement (WJ-R (ach): 42.02), Perceptual and Perceptual-Motor (Bender VMGT: 52.31), Behavior-Social-Emotional (teacher interview: 60.31), Adaptive Behavior (Vineland ABS: 16.60), and Preschool Functioning (Battelle: 23.58). Direct observation of behavior is a task which a majority of psychologists perform often, and which is viewed as being important when assessing the needs of students.

Shapiro (1987) has written that more than any other method of behavioral assessment, systematic direct observation is the most direct and desired approach to the collection of data. Indirect measures of behavior (i.e., teacher and parent interviews, and anecdotal observations) require inferences to be made concerning the reports of informants and the student's actual performance (Cooper, 1987). In contrast, because the data from a direct observation are collected while the behaviors occur, the observation is empirically verifiable and does not necessitate inferences to other behaviors.

Hintze & Shapiro (1996) have identified three advantages of systematic observation: (1) It focuses on the observable behavior of the student. This clear operationalization of the problem makes it possible to clearly identify target behaviors to change. (2) This type of assessment gives the psychologist an opportunity for direct consultation with the teacher. This places the psychologist in a position to draw upon additional skills and provides them with the much preferred role of consultant. And, (3) direct observation provides opportunities for accountability. Legal precedents such as Larry P., PASE v. Hannon, and others remind psychologists that they must assume responsibility for their actions.

As noted earlier, the role of the school psychologist is slowly moving toward service delivery through a number of different avenues. Some have stated that standardized testing could be complimented or even replaced with alternative assessment including curriculum-based measurement (CBM) (Deno, 1985; Germann & Tindal, 1985; Goh & Fuller, 1983; and Ramange, 1979). Alternative methods have their shortcomings, however. Relevant to academic performance is the within-subject variable of cognitive functioning which CBM is reported to ignore (Kamphaus, 1993). In addition, it has been noted that no validation exists for making eligibility decisions using CBM (Shinn et al., 1989).

Although current political and economic trends are directing the profession toward a more expanded role, classification and placement assessments are likely to play an important part, regardless of other activities psychologists will be called upon to perform. Presently, a documented observation is a part of a set of requirements that are needed to classify and place individuals into certain funding categories (e.g., WAC 392-172-128-(3)). These observations are accomplished with varying degrees of depth and complexity. Regardless of how observations are done, the observer spends valuable time calculating and analyzing the data, and writing up a summary of results-all of which often end up in a file. The idea for the Systematic Observation Software (SOS) came from school psychologists who regularly do direct observation. It was thought that a simpler system on a mobile device (i.e., a Personal Digital Assistant (PDA), such as Apple's Newton MessagePad), would prove to be a valuable tool to make this task easier and more efficient.

The information provided by this program will be given to teachers so that specific behavioral objectives may be discussed and new interventions attempted. In addition, the data gleaned from each observation can be uploaded into a database and further analyzed for trends. Baseline data will give teachers and psychologists a foundation from which change and effectiveness of interventions are measured. As the traditional role of the school psychologist shifts, and they are needed to spend more time in regular education classrooms, additional tools and skills will be necessary. Using a problem solving approach and linking assessment to intervention is an important step in this direction. Direct and systematic assessment of behavior is a critical tool in assisting the psychologist in consulting with classroom teachers to aid in the measurement of performance and the implementation of interventions for students. SOS will make direct observation a fundamental part of the full assessment of students.

SOS is not the first attempt to automate the important task of observation. Others have developed systems using computer-based apparatuses to observe subjects in a variety of settings. For example, computers have been used by occupational therapists to observe client behavior to assess the efficacy of treatment (Schneider, Champoux, & Beinert, 1987). Johnson et al., (1995) developed a computer-based system: Direct Observation Data System (DODS) for monitoring performance of students with disabilities. Results from this research demonstrated that data can be collected unobtrusively in the classroom and in the natural settings of the students (e.g., playground, lunchroom, etc.). In another set of studies, a microcomputer system used for behavioral research was found to "save the researchers time in terms of data summary, analysis, and graphing; that the system included no error inherent in the data collection system; and, that the researchers were not limited by number of variables nor to global analyses of behavior" (Repp, Karsh, Acker, Felce, & Harman, 1989; Repp & Felce, 1990). The stand-alone Observational Data Acquisition Program (ODAP) allows for on-line, computer recording of both frequency and temporal features of behavior. Results showed that "the single stroke data entry method, automatic summarization, and data storage to disk minimized personnel training, labor, and human error" (Hetrick et al., 1991). And finally, Greenwood et al., (1994), developers of the EBASS observation software designed to be used on laptop computers, created an "ecobehavioral approach using three instruments that are widely used in special education research-CISSAR, ESCAPE. and MS-CISSAR." These instruments are complex taxonomies of behavior (student, and teacher), and ecology (activity, task, and structure). In order to become proficient in using this system, it is necessary for the observer undergo a 20 hour tutorial that provides definitions of behavior, and periodic calibration with the computer of videotaped observations.

It is hypothesized that participants with minimal training and background in observation, using SOS on the Newton, will demonstrate the reliability of the instrument. If the instrument is simple enough to be given to an observer who has had no prior experience with either PDAs, nor extensive training on observational techniques, SOS could be utilized by a diverse group of professionals. This pilot study will demonstrate both the development and the interobserver reliability of the interval option of the Systematic Observation System. Kappa, which controls for chance, will be the statistic used to determine the reliability of SOS. (The final version of SOS is currently in the programming and debugging phase. The version that was used to determine reliability for the study was a simplified model of what will be discussed in other parts of this paper.)

METHOD

Ethics Review

The Eastern Washington State University Committee on the Use of Human Subjects in Research has reviewed this study and determined that the rights and welfare of the human subjects were sufficiently protected. The risks were outweighed by the potential benefits and predicted value of the information sought. Confidentiality of data was assured, and informed consent was achieved by appropriate procedures.

Participants

Participants' ages ranged from 25-50, and gender was evenly distributed (8 males and 9 females; N=17). Participants came from a diverse background, and fit into one of three classifications: currently practicing school psychologists (n=7), graduate level clinical psychology students (n=7), and other school practitioners (i.e., consulting teachers, counseling assistants, etc.) (n=3). The range of participants who had formal observation training and experience performing behavior observations was from none to graduate level.

Apparatus

The computer industry has made some dramatic advances in the last 15 years. Moving from desktop computers to laptops, notebooks, and most recently palmtops, information processing machines have become smaller and more efficient. Recently, the computer industry has developed even smaller devices culminating in the idea of the PDA. The Newton, from Apple, Inc., is but one of many PDAs to appear on the market. Unlike all PDAs, however, the Newton has no keyboard, and no connecting mouse. It presents the user with a pen and pressure sensitive screen. This interface between the user and the computer is analogous to the traditional pencil and sheet of paper which has been the standard. Unlike a piece of paper, however, SOS on the Newton is intelligent; Each observation form can be modified by the user. Because school psychologists often travel from building to building, the hand held Newton seemed to be compatible with this profession-the traveling data collector.

SOS, written for Newton's operating system, was designed to help the school psychologist perform an observation with no paperwork and minimal effort in terms of written information, calculation, and analysis of data. (Almost all screens utilize pop-up menus to significantly reduce the amount of writing and simplify the coding of behaviors.) SOS was designed not only to be less cumbersome, but more efficient. Processing at speeds up to 20 megahertz, with 2,000K of RAM, the compact computer quickly tallies and sums rows and columns of data. Also, because data are stored in battery backed up RAM, there is no waiting time for the operating system to boot-up from a hard drive. In other words, when the user presses the "on" switch, the Newton is instantly up and ready for data input.

Systematic Observation Software

Present and future directions of SOS are described and graphically represented below:

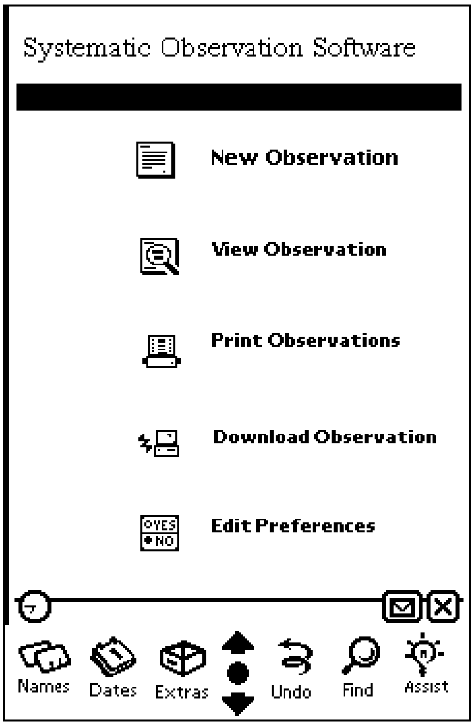

Figure 1

Introduction Screen

From the main start up screen, the user has a number of different choices: New Observation, View Observation, Print Observation, Download/Upload Observation, and Preferences. When the "New Observation" button is pressed, an input screen will ask for the child's teacher's name, the school, and other demographic information. After this information has been entered, the Observation Type Screen appears, and presents the user with the four types of observations: event, interval, narrative, and ratings (see Figure 2).

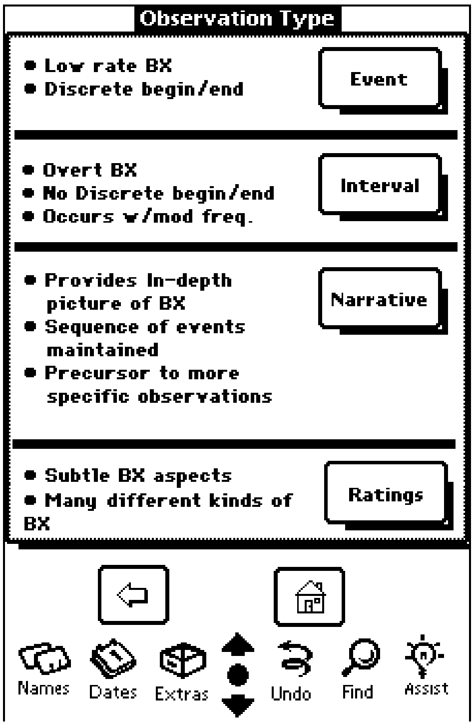

Figure 2

Observation Type Screen

At this point, SOS alerts the observer to the advantages of four types of observations available. Each observation has been developed to assess a variety of behaviors, and each has a different purpose (Shapiro, 1987). For example, if the psychologist is to observe a child who is walking around the room, and the child exibits this behavior occasionally, an event recording would be the best choice. The reason: behaviors that occur intermittently throughout the day and that have a definite beginning and ending are best observed by counting the number of times the event occurs. However, if that same child were daydreaming intermittently throughout the day (a behavior that does not have a discrete beginning and end, is overt, and occurs moderately), counting every instance might prove to be difficult. Observing the student for a few seconds, in intervals of 30 seconds, is a more manageable task.

After the user has decided that he or she will use the interval recording method, for example, a screen will appear and instruct the user to set the timing, frequency, type, and other necessary information for the Newton to cycle automatically (Figure 3).

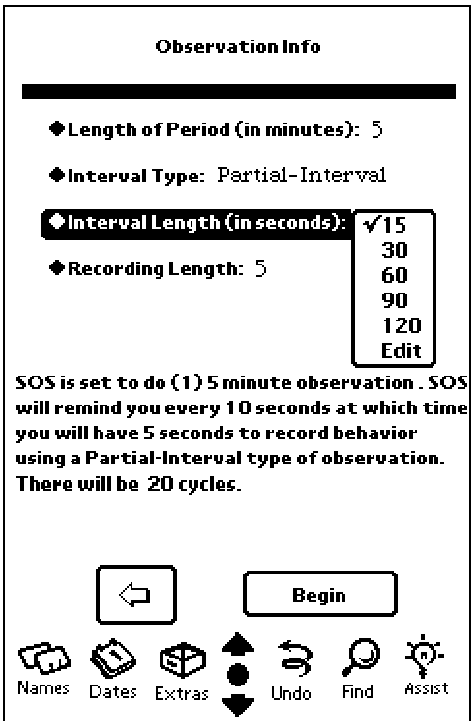

Figure 3

Observation Information Screen

The Observation Information screen contains the following choices: number of observations, length of period, interval type, interval length, and recording length. By simply tapping on the black diamond to the left of the phrase, a menu automatically pops up, waiting for the user to pick a number. The paragraph displayed under the options is a summary of what the user has chosen. When the "Begin" button is pressed, SOS begins cycling, and the Data Input screen appears (see Figure 4).

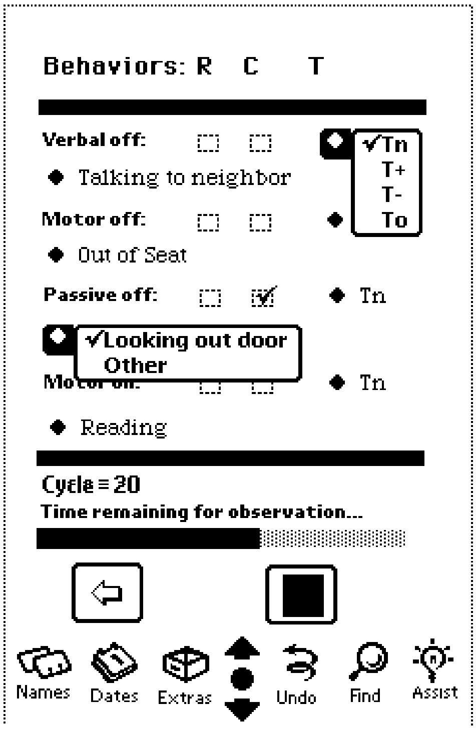

Figure 4

Data Input Screen (Interval)

This screen allows the observer to press a button to indicate an occurrence of the referred child's target behavior(s) and the teacher's reaction, and a comparison child's behavior and teacher's reaction. "R" represents the referred child, "C" stands for a randomly chosen comparison child, a child to compare to the referred child, and "T" is for the teacher or third party. The pop-up options represent his/her reactions to the other children's behavior (Tn=a neutral response by the teacher, T+=teacher approval, T-=disapproval, TO=no response). Over time, the responses of the teacher will be paired with the responses of the children. In this way the observer will be able to make some comparisons, and recommendations. The program visually (blinking screen) and aurally (paper turn, writing sound, beep, etc.) reminds the user when to start observing and stop observing at each interval. Another feature is the ability to preload specific sentences that describe exactly what he or she sees (e.g., not only "Passive Off," but, "Looking out door," and other ways of defining what the user means by, "Passive Off"). When an unplanned event occurs (e.g., a fire drill, a rock hits the window, etc.), or other events that occur before the target behaviors, the user can press the Antecedent ("A") button to record any anecdotal information. This option allows the psychologist to use his or her clinical judgment, and to add antecedent information vital to making intervention recommendations.

Information from the previous screen is analyzed, and data placed on a Results/Summary Analysis screen. The Results/Summary Analysis screen will include: a total count of the highest occurring behaviors (verbal/motor/passive on- and off-task) for both referred and comparison students, a breakdown and total count of specific behaviors that occurred from the previous six categories (using the pop-up phrases), rates of behaviors for both students, and a comparison of the teacher's reactions to both students. In addition to the results of the observation, SOS will use the previous data, and other information (e.g., name, location, teacher, and other relevant demographics) from the Information screen to produce a narrative account of the observation. This narrative report will examine the antecedents and consequences of behaviors for both students in addition to information regarding the location, time of day and other environmental factors. The user will have the ability to modify this report, SOS simply provides a framework into which the psychologist can integrate experience and judgment.

As stated earlier, the version of the program that was used for the pilot study was a simplified version of what is presented above. For this study, an interval observation was the only option available (See Figures 5 and 6). Passive, motor, and verbal on- and off-task behaviors were displayed without pop-up menus. Interval recording is the most common type of behavior observation (Sattler, 1992), and was determined to be the clearest method of collecting quantifiable and comparable data to determine the reliability of the interval option on SOS.

From its inception, SOS evolved into a program that would not only tally behaviors, but a powerful tool that could be configured to suit the purposes of the situation and psychologist. If, for example, the observer wanted to first obtain global data on a student, he or she could simply choose the narrative option on SOS and a list of questions that an experienced observer would ask him or herself is displayed. After handwriting the information directly into the computer, the observer has the option of uploading that observation into a desktop computer and adding it to the file of that child, or printing it out directly from the Newton to discuss with the teacher.

When more quantitative data are required, the observer simply decides the type of observation that the situation dictates (e.g., interval, event, or ratings). From this point, the following information is needed: (1) the number of people to be observed (e.g., a referred child in addition to a comparison child and/or a teacher's reactions), (2) the target behaviors to be recorded, (3) the time and place, and (4) other information needed for the report generated by SOS. In the Event Observation mode, for example, SOS waits for input and informs the observer when the time for the observation has elapsed. The Results Screen displays a breakdown and summation of target behaviors, rates of behavior for the referred child (and comparison child, if chosen as an option), and a chronological account of recorded target behaviors.

When the Interval Observation option is selected, SOS guides the observer through the timing sequence and awaits a push of the "Begin" button. Like the Event option, Interval observations allow the user to conform the observation to fit the situation. All the pop-up screens are equipped with an edit function which permits the user to change and manipulate data for each observation. This allows the observer to fit the observation to the specific classroom. SOS is unique in its ability to both time the intervals, and allow the observer to select pre-written phrases all with the tap of a pen. For example, if a child is recorded as being motor off-task, a number of different conclusions could be drawn. If, however, the user wanted to obtain both quantitative and qualitative information, the observer simply checks the box next to the broad category of behavior (i.e., verbal/motor/passive on, and off-task), then taps the diamond under that category to see a list of phrases that describe precisely what is happening (e.g., motor off-task behavior could include the following list of behaviors or phrases: "out of seat," "throwing objects," "spitting," "aggression," "head banging," or any other sentence that the psychologist typically uses). This type of observation is unique to SOS, and it is made possible because of Newton's innovative design.

Procedure

Participants were asked to view a 25 minute video of a fifth-grade, male child in a regular education classroom. The video used was from the EBASS calibration tutorial (see Appendices A and B for various letters of permission). Participants recorded verbal/motor/passive on-task and off-task behaviors in intervals of 20 seconds-10 seconds to observe and 10 seconds to record. The Data Input Screen was displayed for 10 seconds (see Figure 5), allowing the observer to code the behavior. The "Observe" screen (see Figure 6) would then appear for 10 seconds locking out all data entry. This assured that the appropriate time had passed for that interval.

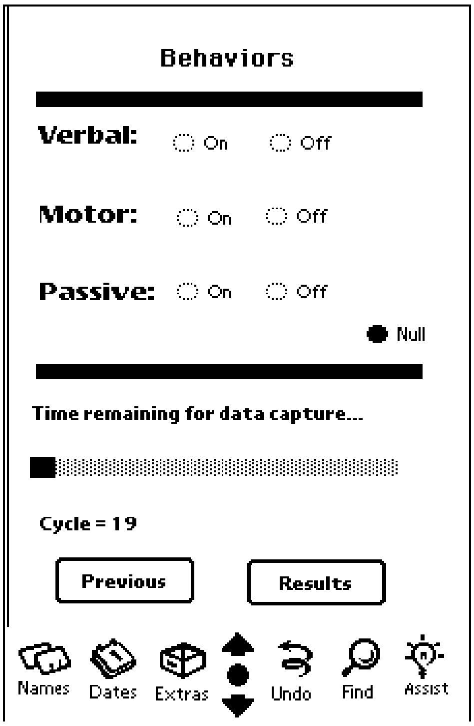

Figure 5

Data Input Screen (Interval-simplified)

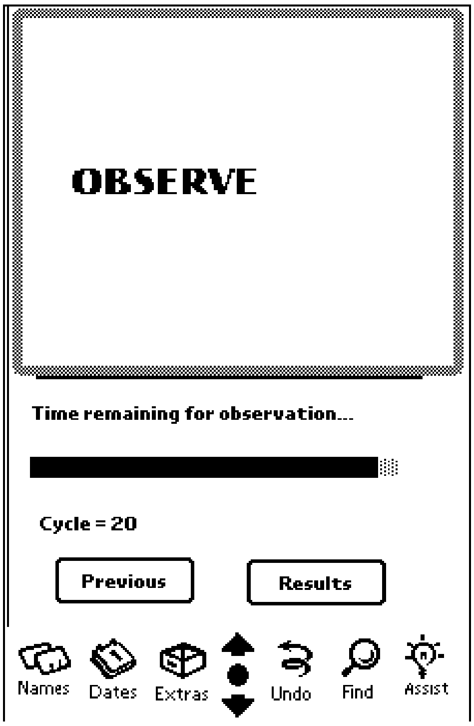

Figure 6

Observe Screen (Interval-simplified)

Four activities were observed:

Activity I: Science; Group instruction and group work (two students)

Duration: 11:20

Rules: **Listen to teacher**

Work in group

Reader: oral reading of text

Listener: write responses on paper

Activity II: Spelling; One on one instruction

Duration: 4:45

Rules: **Work alone**

Practice with SPED teacher

Activity III: Reading; Oral reading to teacher/class; Lecture

Duration: 5:25

Rules: **Get out book**

Open book

Read chapter as teacher directs

Activity IV: Math; Group instruction

Duration: 3:40

Rules: **Remain at board/desk**

Write answers to teacher's questions on board/paper

A whole-interval type of recording procedure was used in this study. This meant that on-task behavior was scored only when it occurred at the beginning of the interval and lasted throughout the entire interval. All intervals that contained even one off-task behavior were coded as off-task. In the few cases that more than one off-task behavior was demonstrated, the participants were instructed to record the behavior that occurred first. It was explained to the participants that any break in eye or manual contact with the task materials resulted in that interval's being scored as off-task. The participants were also told that at times the camera panned away from the target child, when this happened, they were to record responses based on what they heard. It was explained that if this were not possible, they were to record the behavior that was seen last.

Participants were given a handout that described on- and off-task behavior (See Appendix B). They were directed to refer back to this paper that described the activity and criteria for on-task behavior as time between intervals allowed. After the brief training, the participants were left alone in a room equipped with a television, video recorder, and Newton. Participants were checked once after 10 minutes of observation to determine the status of the equipment. In addition to the verbal/motor/passive off- and on-task behavior classifications, the following clarification of behavior was explained:

Verbal on: **appropriate vocalization** in the context of current rules in effect

Verbal off: **inappropriate vocalization** in the context of current rules in effect

Motor on: **appropriate movement** in the context of current rules in effect

Motor off: **inappropriate movement** in the context of current rules in effect

Passive on: **apparent concentration** in the context of current rules in effect

Passive off: **apparent daydreaming** in the context of current rules in effect

On-task behavior must have met the following criteria:

(a) the target child's buttocks must be in the appropriate place in the context of the activity while

(b) his eyes must be oriented to the task materials while

(c) he interacts manually with the task materials.

All participants' responses were correlated with the combined expert opinions of several school psychologists who had been practicing observations for 15 or more years. The total number of observed intervals was 75. Because participants were given only a few minutes to become comfortable with this instrument, their responses for the first five minutes were not included in the data. In addition, in order to control for the abrupt changes from activity to activity, the pre and post interval data for the transitions were not included in the results.

RESULTS

Results indicate an overall acceptable level of interobserver reliability (kappa: M=.73; Range=.67-.83) when the invalid data from

Activity IV are removed (see Table 1). Participants classified as "Other School Personnel" demonstrated the highest reliability ranging individually from .79 to .85 (M=.83). The second highest group of participants were the graduate level students (M=.70), and third, practicing school psychologists (.67) (See Appendices D-F for raw data).

Table 1

Reliability Coefficients for Three Groups of Participants

	Activity I	Activity II	Activity III	Activity IV	Total
O	.78	.81	.89	--	.83
S	.57	.82	.72	--	.70
P	.46	.85	.70	--	.67
	.60	.83	.77	--	.73

O=Other School Personnel;

S=Clinical Psychology Graduate Students;

P=School Psychologists.

ACTIVITY I=Science: Group instruction and group work (two students);

ACTIVITY II=Spelling: One on one instruction;

ACTIVITY III=Reading: Oral reading to teacher/class; Lecture;

ACTIVITY IV=Math: Group instruction.

Kappa is a statistic that controls for chance agreement. When kappa is positive, the proportion of observed agreement is more than the proportion of chance agreement. When kappa is negative, the proportion of observed agreement is less than the proportion of chance agreement. A standard acceptable level of kappa is a positive integer equal to or greater than .70. (See Appendix G for a detailed comparison of kappa and total percentage agreement).

Reliability for the first observed activity was unacceptable (M=.60), while activities two and three were found to be acceptable (M=.83, M=.77). Because not all participants produced an equal number of total observed intervals, the data for the last activity were incomplete and therefore incomparable. The data for Activity IV were consequently omitted from the results.

DISCUSSION

The three activities that were used to determine the reliability of SOS were as follows: Science: Group instruction and group work (two students) (ACTIVITY I), Spelling: One on one instruction (ACTIVITY II), Reading: Oral reading to teacher/class; Lecture (ACTIVITY III). The total number of observed and analyzed intervals was 44 (excluding the 10 minute training period and invalid Activity IV data). Of the total number of intervals, 50% were to be coded as being on-task, and 50% off-task (22 intervals on, 22 intervals off).

The highest reliability coefficient that was obtained was for the spelling activity (.83). This activity consisted of 13 observed intervals of time (approximately 4 min. 30 sec.) During this segment of the video, the target child remained at a desk while a special education teacher sat in front of him, helping him review his spelling words. On-task intervals accounted for 77% of the total time (23% off-task).

The second highest reliability coefficient was found for oral reading (.77). This activity consisted of 14 intervals of time (5 min. 30 sec.). The target child was to first remove his book from his desk, then open it, and, when asked, to orally read parts of a text both to the teacher and to the class. On-task intervals accounted for 36% of the total time (64% off-task).

The third highest reliability coefficient was demonstrated for the Science activity (.60). Activity I consisted of 32 intervals of time which translates into approximately 11 min. and 30 sec. (analyzed data accounted for approximately 17 intervals: 6 min. 30 sec.). The activity began with group instruction and then the students were broken into groups of two. One member orally read the chapter, and the other wrote answers to questions. On-task intervals accounted for 41% of the total time (59% off-task). As stated earlier, the results for Activity IV were found to be invalid because of incomplete and chronologically mismatching data.

Activity I resulted in the lowest and only unacceptable level of interobserver reliability (.60). A number of hypotheses for this finding can be made: (1) the lack of experience of the participants with the PDA and/or SOS, (2) the ambiguous nature of the child's behavior in the context of the rules in effect, and (3) the timing discrepancy between the video and software. What follows are additional explanations of the sources of unreliability. Cronbach (1970) developed a framework for examining variation as it relates to reliability. The two factors that make up a two-by-two table are temporal and generality (see Table 2). Each factor is then broken down into two levels: generality consists of specific and general effects, while the temporal factor consists of temporary and lasting effects. Specific effects are those sources related to the specific instrument being used, and general effects include sources of variation that influence the reliability of any instrument. Temporary effects are those that influence reliability for a short time, whereas lasting effects persistently affect reliability over time.

Table 2

Factors affecting the reliability of observation instruments

TEMPORAL FACTORS

Lasting Temporary

Specific Knowledge of coding scheme Attitude toward activities being observed Observation constraints Reading wiseness	Fatigue due to current task Attention span Memory of event Memory of coding scheme Practice effect of coding task Luck or guessing
General Intellectual ability Skill in observing Observer-wiseness	Observer drift Health, fatigue, emotions Motivation

Attitude toward reading Molar or molecular level of observation task Hawthorne effect	Physical surroundings Amount of practice

Note. From "Practical suggestions for increasing the reliability of classroom observational research," Mangano et al., 1986, Reading , Research, and Instruction, 25 (3), p.186

Lasting-Specific Sources

Lasting-specific sources of variation affecting the reliability of SOS relate to the observer's familiarity with the coding system, attitudes, and training (Mangano, Willson, & Rupley, 1986). These were controlled for in a number of ways including simplifying the coding system (using "radio buttons" to indicate the occurrence of a behavior). From its inception SOS was created to be a user-friendly device, and so training given to participants was brief (approximately 5 minutes) and simple (see Appendix C for written instructions to participants). Although all of the participants had some fundamental understanding of observation, none had any experience in using SOS or the Newton.

Lasting-General Sources

SOSs lasting-general factors to consider were the cognitive ability of observers, observation skills, complexity of the observation and supervision conditions (Mangano et al., 1986). The three categories of participants were school psychologists, graduate students, and other school personnel. All participants were aware of the purpose of this study, and all seemed to be capable and competent. The participants were aware that the data would be collapsed into on- v. off-task behaviors. This was done to simplify the observation and to ease the stress of the participants. Frequent supervision has been shown to influence the reliability of the observer (Kazdin, 1977), however, this study attempted to examine a situation that would represent a realistic observation setting.

Temporary-Specific Sources

These sources of unreliability are often outside the researcher's control. They include factors like fatigue, attention span, memory of events, practice effects, and guessing (Mangano et al., 1986). The total amount of unaccompanied observation time required of the participants was 15 minutes. This is 5-10 minutes less than what is typically spent doing an observations in a classroom. Also, because of the scoring procedures (i.e., if any off-task behavior was seen, the entire interval was recorded as off-task) memory of events should not have played a significant role in determining the reliability of SOS.

Temporary-General Sources

Temporary-general sources of variations which affect all systems, but are of limited duration, are factors such as physical location and surroundings, observer drift, and the physical condition of the observer (Mangano et al., 1986). Participants were able to perform the observation in a secluded setting, and distractions were reduced as much as possible. Observer drift refers to observers changing the definition of behaviors over time. This might have been a factor in explaining the unreliability of earlier activities. It has been found that by allowing the observer to code behaviors for a set period of time before actually gathering, the accuracy of the data is improved (Mangano et al., 1986). This technique was included in this study.

Summary

School psychologists perform many direct observations of students' behavior. As school psychology evolves into a profession concerned more with functional assessment rather than administrative assessment, observations will become an important tool in the expanding repertoire of psychologists. Through this pilot study, it was found that the interval version of Systematic Observation Software developed for the Newton, possesses an acceptable level of reliability.

Two limitations of this pilot study are: (1) the relatively low number of participants, and (2) no attempts to ascertain validity. Suggestions for further research might include more training of participants- including other types of participants in the sample who might be doing observations (i.e., teachers, administrators, other special education personnel, MSWs, behavioral psychologists, etc.), increasing the number of participants in a replication study, attempting to standardize and provide some normative data, and attempting to assess the validity of results.

Also, future versions of SOS will contain features that might increase its effectiveness as a research tool. Three examples are: (1) using the "Beam" function, the Newton could be programed to start the video recorder. This improvement would address the possible factor of unreliability associated with the video-computer synchronization. (2) Add a function that would enable the observers to determine interobserver reliability either during or after an observation. (By "Beaming" results to a partner, the Newton would calculate the percentage agreement for the observation up to that moment.) And, (3) program SOS to perform simple statistics to determine significant differences of observed behaviors (e.g., a t-test to examine deviance of a referred child from a comparison child).

REFERENCES

Alessi, G.J. (1980). Behavioral observation for the school psychologist: Responsive-discrepancy model. School Psychology Review, 9 (1), 31-45.

Batsche, G.M. (1992). School Psychologists and assessment: A future together? Communique, p. 2

Cooper, J.O. (1987). Measuring and recording behavior. In J.O. Cooper, T.E. Heron, & W.L. Heward (Eds.), Applied behavior analysis. Columbus: Merrill Publishing.

Cronbach, L.J. (1970). Essentials of psychological testing. New York: Harper & Row.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Germann, G., & Tindal, G. (1985). An application of curriculum-based measurement: The use of direct and repeated measurement. Exceptional Children, 52, 244-265.

Goh, D.S., & Fuller, G. B. (1983). Current practices in the assessment of personality and behavior by school psychologists. School Psychology Review, 12, 240-243.

Graden, J. L., Zins, J.E., & Curtis, M.J. (1988). Alternative educational delivery systems: Enhancing instructional options for all students. Washington: National Association of School Psychologists.

Greenwood, C.R., Carta, J.J., Kamps, D., Terry, B., Delquadri, J. (1994). Development and validation of standard classroom observation systems for school practitioners: Ecobehavioral assessment systems software (EBASS). Exceptional Children, 61 (2), 197-210.

Hetrick , W.P., Isenhart, R.C., & Taylor, D.V. (1991). ODAP: A stand alone program for observational data acquisition. Behavior Research Methods, Instruments, and Computers, 23 (1), 66-71.

Hintze, J.M., & Shapiro, E.S. (1996). Best practices in the systematic observation of classroom behavior. Thomas, Alex (Editor); Grimes, Jeff (Editor) Best Practices in School Psychology III: National Association of School Psychologists.

Johnson, H., Blackhurst A. E., Maley, K., Bomba, C., Cox-Cruey, T., & Dell, A. (1995). Development of a computer-based system for the unobtrusive collection of direct observational data. Journal of Special Education Technology, 7 (4), 291-300.

Kamphaus, R.W. (1993). Clinical Assessment of Children's Intelligence. Boston: Allyn & Bacon.

Kazdin, A.E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. Journal of Applied Behavioral Analysis, 10, 141-150.

Lynch, W. W. (1977). Guidelines to the use of classroom observation instruments by school psychologists. School Psychology Monograph, 3 (1), 1-22.

Mangano, N.G., Willson, V.L., & Rupley, W.H. (1986). Practical suggestions for increasing the reliability of classroom observational research. Reading Research and Instruction, 25 (3), 184-191.

Naglieri, J.A., Das, J.P., & Jarman, R.F. (1990). Planning, attention, simultaneous and successive cognitive processes as a model for assessment. School Psychology Review, 14, 423-442.

Ramange, J.C. (1979). National survey of school psychologists: Update. School Psychology Digest, 8, 153-161.

Repp, A.C., Karsh, K.G., Acker, R.V., Felce, D., & Harman, M. (1989). A computer-based system for collecting and analyzing observational data. Journal of Special Education Technology, 9 (4), 207-217.

Repp, A.C., & Felce, D. (1990). A microcomputer system used for evaluative and experimental behavioral research in mental handicap. Mental Handicap Research, 3 (1), 21-32.

Reschly, D.J. (1980). School psychologists and assessment in the future. Professional Psychology, 11, 841-848.

Reschly, D.J. (1988). Special education reform: School psychology revolution. School Psychology Review, 17, 450-475.

Rowley, G. L. (1988-89). Reliability and other indicators of data quality in classroom observation research. Journal of Classroom Interaction, 24 (1), 22-29.

Sattler, J.M. (1992). Assessment of Children: Revised and Updated Third Edition. San Diego: Jerome M. Sattler, Publisher, Inc.

Schneider, M. L., Champoux, M., & Beinert, R. H. (1987). The use of computer-assisted behavior observation in sensory integration practice and research. Occupational Therapy in Health Care, 4 (2), 129-136.

Shapiro, E.S. (1987). Behavioral Assessment in School Psychology. Lawrence Erlbaum Associates, Incorporated.

Shinn, M.R., Rosenfield, S., & Knutson, N. (1989). Curriculum-based assessment: A comparison of models. School Psychology Review, 18, 299-316.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. Journal of Psychoeducational Assessment, 12 (4), 331-350.

APPENDICES

Appendix A

Thaynan Knowlton

19816 S. Culver Rd.

Cheney, WA 99004

(509) 235-1615

April 12, 1996

Dr. Charles Greenwood

1614 Washington Blvd.

Kansas City, KS 66102

Dear Dr. Greenwood,

I thank you for your prompt response to me via the telephone for your permission to use the regular education calibration tape for EBASS. I would like to use it in a interrater reliability study on a new instrument for my Master's thesis. (An abbreviated copy of my formal proposal to the ethics committee is attached.)

I have another request. Will you send me a letter giving me permission to proceed?

Thank you for your time,

Thaynan Knowlton

DEVELOPMENT AND RELIABILITY
OF AN OBSERVATION SYSTEM:
PERSONAL DIGITAL ASSISTANT (PDA)
SYSTEMATIC OBSERVATION SOFTWARE (SOS)

A. Rationale for this study:

Special education law requires a documented observation of a child as part of a set of requirements to classify children into funding categories. Because a high percentage of children fall into the Specific Learning Disability category, for example, valuable time is spent observing children directly in the classroom, calculating and analyzing the data, and writing up the summary of information to place in a file for a child. Currently, many school practitioners use paper and pencil observation forms that are either too simplistic or overly complicated and time consuming. Also, the information from these observations is rarely used in interventions designed to directly and immediately benefit the referred child.

Recently, computer miniaturization has become dramatic, leading to the concept of the Personal Digital Assistant (PDA), with pen/touch based interfaces. The Systematic Observation Software (SOS) written for Apple's PDA, "The Newton," will allow school practitioners to observe the classroom without distracting paperwork and timing devices. It also requires minimal effort in terms of written information and the calculation and analysis of data. The information provided by SOS can be given to the teacher so that specific behavioral objectives may be discussed and new interventions attempted. Baseline data will give teachers and psychologists a foundation from which to measure change and effectiveness of interventions. In addition, the data gleaned from each observation could be uploaded into a database and analyzed for trends. Although other computer observation software exists (i.e., EBASS, DODS) for laptop computers and other PDAs, SOS is unique in that it is specifically designed for the Newton.

B. Objectives of this specific research:

The primary focus of this project is the development of SOS by the author and a programmer, and by practicing school practitioners field testing the software. For this specific study however, 15 minutes of videotaped classroom activity will be shown to practicing school psychologists, and school psychology students to determine the interobserver reliability of the instrument. Participants will use SOS to observe and record behavior in intervals of 30 seconds (20 seconds observation interval, 10 seconds recording interval) for a child in the classroom. The participants will observe motor/verbal/passive off-task and motor/verbal/passive on-task behaviors. A brief training session for participants will take place prior to the trial explaining the definitions of the specific target behaviors, and demonstrating the use of SOS and the Newton. Interobserver reliability will be demonstrated using Kappa.

C. How subjects will be involved and what they will do:

Psychologists will observe videotaped classroom activity using SOS. Students will be engaged in normal instruction and will not be nominally identified.

D. How will obtained data answer the research problem:

By correlating observed responses of psychologists, the research will demonstrate that SOS has reasonable interrater reliability. Implications for these findings are that the software can be used to reliably measure off-, and on-task behaviors. This, in turn, can be used to identify and measure specific interventions.

Appendix B

PERMISSION FORM

DEVELOPMENT AND RELIABILITY OF AN OBSERVATION SYSTEM: PERSONAL DIGITAL ASSISTANT (PDA) SYSTEMATIC OBSERVATION SOFTWARE (SOS)

P.I.: Thaynan Knowlton, Graduate Student, Psychology Department, 235-1615

PURPOSE AND BENEFITS

By demonstrating that SOS is a reliable instrument, school personnel working in a school district will be able to quickly and simply conduct systematic, direct observation of children. Information gathered from these observations can be used in interventions designed to directly and immediately benefit the referred child.

In addition, the information provided by SOS can be given to the teacher so that specific behavioral objectives may be discussed and new interventions attempted. Baseline data will give teachers and school personnel a foundation from which to measure change and effectiveness of interventions. Also, the data gleaned from each observation could be put into a database and later analyzed for trends.

PROCEDURES

For this specific study, approximately 15 minutes of videotaped classroom activity from the Ecobehavioral Assessment Systems Software (EBASS) calibration regular education videotape will be shown to qualified school personnel to determine the interobserver reliability of the instrument. Participants will use SOS to observe and record behavior in intervals of 30 seconds (20 seconds observation interval, 10 seconds recording interval) for a child in the classroom. The participants will observe motor/verbal/passive off-task and motor/verbal/passive on-task behaviors. A brief training session for participants will take place prior to the trial explaining the definitions of the specific target behaviors, and demonstrating the use of SOS and the Newton. Inter-observer reliability will be demonstrated using kappa. Total time involved will be between 30 and 60 minutes, including set-up and take-down time. There will be no questionnaire or interview.

RISK, STRESS OR DISCOMFORT

There is little or no foreseeable risk for psychological or physical harm in this study. The study will involve nothing more than what people typically experience on a day to day basis. Also, participant's recorded responses will not be identified nor linked to them.

OTHER INFORMATION

Your school personnel's identity will remain confidential, and he or she is free to withdraw anytime without penalty. The study described above has been explained to me, I have had an opportunity to ask questions and I give my permission for the appropriate school personnel under my jurisdiction to participate in this activity.

____

Signature of School Administrator Date

Appendix C

DEVELOPMENT AND RELIABILITY OF AN OBSERVATION SYSTEM: PERSONAL DIGITAL ASSISTANT (PDA) SYSTEMATIC OBSERVATION SOFTWARE (SOS)

EASTERN WASHINGTON UNIVERSITY

PRINCIPLE INVESTIGATOR: THAYNAN KNOWLTON, GRADUATE STUDENT, DEPARTMENT OF PSYCHOLOGY

RATIONALE:

Special education law requires a documented observation of a child as part of a set of requirements to classify children into funding categories. Because many children fall into the Specific Learning Disability category, for example, valuable time is spent observing children directly in the classroom, calculating and analyzing the data, and writing up the summary of information to place in a file for a child.

SYSTEMATIC OBSERVATION SOFTWARE (SOS) FOR THE NEWTON:

Recently, computer miniaturization has become dramatic, leading to the concept of the Personal Digital Assistant (PDA), with pen/touch based interfaces. The Systematic Observation Software (SOS) written for Apple's PDA, Newton, will allow school practitioners to observe the classroom without distracting paperwork and timing devices. It also requires minimal effort in terms of written information and the calculation and analysis of data. The information provided by SOS can be given to the teacher so that specific behavioral objectives may be discussed and new interventions attempted. Baseline data will give teachers and psychologists a foundation from which to measure change and effectiveness of interventions. In addition, the data gleaned from each observation could be uploaded into a database and analyzed for trends. Although other computer observation software exists (i.e., EBASS, DODS) for laptop computers and other PDAs, SOS is unique in that it is specifically designed for the Newton.

INTEROBSERVER RELIABILITY STUDY:

You are about to see approximately 25 minutes of classroom activity of a fifth grade regular education classroom. You will record verbal/motor/passive on-task and off-task behaviors in intervals of 20 seconds-10 seconds to observe and 10 seconds to record. Because you will be given only a few minutes to become comfortable with this instrument, your responses for the first five minutes will not be included in the data. Your responses are confidential and the printed data will be aggregated so that your identity will be anonymous.

INSTRUCTIONS:

Below are the four activities that you will see during the next 25 minutes. Please read and take note of the rules in effect for each activity as this will influence how you record the behavior of the target child. A whole-interval type of recording procedure will be used in this study. This means that on-task behavior is scored only when it occurs at the beginning of the interval and lasts throughout the entire interval. All intervals that contain even one off-task behavior should be coded as off-task. In the case that more than one off-task behavior is demonstrated, record the behavior that occurred first. In addition to the classifications below, on-task behavior must meet the following criteria: (a) the target child's buttocks must be in the appropriate place in the context of the activity (b) his eyes must be oriented to the task materials while (c) he interacts manually with the task materials. Any break in eye or manual contact with the task materials results in that interval's being scored as off-task-the first off-task behavior demonstrated. At times the camera will pan away from the target child, when this happens, please record responses based on what you hear. If this is not possible, record the behavior that you last saw.

The following classifications will be used to further describe behavior:

Verbal on: **appropriate vocalization** in the context of current rules in effect

Verbal off: **inappropriate vocalization** in the context of current rules in effect

Motor on: **appropriate movement** in the context of current rules in effect

Motor off: **inappropriate movement** in the context of current rules in effect

The following are reserved for special circumstances:

Passive on: **apparent concentration** in the context of current rules in effect

Passive off: **apparent daydreaming** in the context of current rules in effect

Activity I: Science; Group instruction and group work (two students)

Time: 11:20

Rules: **Listen to teacher**

Work in group

Reader: oral reading of text

Listener: write responses on paper

Criteria: (a) the target child's buttocks must be touching the seat of his chair,

(b) his eyes must be oriented to the task materials while

(c) he interacts manually with the task materials.

Activity II: Spelling; One on one instruction

Time: 4:45

Rules: **Work alone**

Practice with SPED teacher

Criteria: (a) the target child's buttocks must be touching the seat of his chair,

(b) his eyes must be oriented to the task materials while

(c) he interacts manually with the task materials.

Activity III: Reading; Oral reading to teacher/class; Lecture

Time: 5:25

Rules: **Get out book**

Open book

Read chapter as teacher directs

Criteria: (a) the target child's buttocks must be touching the seat of his chair,

(b) his eyes must be oriented to the task materials while

(c) he interacts manually with the task materials.

Activity IV: Math; Group instruction

Time: 3:40

Rules: **Remain at board/desk**

Write answers to teacher's questions on board/paper

Criteria: (a) the target child's feet must be in his assigned area,

(b) his eyes must be oriented to the task materials while

(c) he interacts manually with the task materials.

You may refer back to this sheet during any extra recording-interval time.

Appendix D

OTHER SCHOOL PERSONNEL:

	Activity I	Activity II	Activity III	Activity IV	Total
O1	.72	.81	1.0	--	.84
O2	.72	.81	.84	--	.79
O3	.91	.81	.84	--	.85
	.78	.81	.89	--	.83

O=Other School Personnel. ACTIVITY I=Science: Group instruction and group work (two students); ACTIVITY II=Spelling: One on one instruction; ACTIVITY III=Reading: Oral reading to teacher/class; Lecture; ACTIVITY IV=Math: Group instruction.

Appendix E

GRADUATE STUDENTS (MS CLINICAL PSYCHOLOGY):

	Activity I	Activity II	Activity III	Activity IV	Total
S1	.60	1.0	.69	--	.76
S2	.53	.75	.66	--	.65
S3	.30	1.0	.38	--	.56
S4	.76	1.0	1.0	--	.92
S5	.61	.41	.46	--	.49
S6	.29	.56	.85	--	.57
S7	.88	1.0	1.0	--	.96
	.57	.82	.72	--	.70

S=Clinical Psychology Graduate Students. ACTIVITY I=Science: Group instruction and group work (two students); ACTIVITY II=Spelling: One on one instruction; ACTIVITY III=Reading: Oral reading to teacher/class; Lecture; ACTIVITY IV=Math: Group instruction.

Appendix F

SCHOOL PSYCHOLOGISTS:

	Activity I	Activity II	Activity III	Activity IV	Total
P_1	.53	1.0	.38	--	.64
P_2	.77	.42	.51	--	.57
P_3	.39	.75	.66	--	.60
P_4	.33	1.0	.84	--	.72
P_5	.13	1.0	.67	--	.60
P_6	.88	.81	.84	--	.84
P_7	.20	1.0	1.0	--	.73
	.46	.85	.70	--	.67

P=School Psychologists. ACTIVITY I=Science: Group instruction and group work (two students); ACTIVITY II=Spelling: One on one instruction; ACTIVITY III=Reading: Oral reading to teacher/class; Lecture; ACTIVITY IV=Math: Group instruction.

Appendix G
Comparison of Two Different Observer Agreement Formulas

		O_2	
		O	NO
O_1	O	a	b
	NO	c	d

Total

Percentage

Agreement Kappa

1. 100 .00

50	0
0	0

2. 98 .00

49	1
0	0

3. 98 .66

48	1
0	1

4. 98 .00

0	1
0	49

5. 100 1.00

25	0
0	25

6. 98 .96

24	1
0	25

7. 80 .62

20	10
0	20

8. 50 .00

9	16
9	16

9. 4 -.92

24	1

1	24

10. 50 .00

14	12
13	11

O1=Observer 1, O2=Observer 2, O=occurrence, NO=nonoccurrece

- Back