Running Head: SKILL AND PERFORMANCE DEFICITS

Identifying Academic Skill and Performance Deficits:

The Experimental Analysis of Brief Assessments of Academic Skills

Gary J. Duhon

Oklahoma State University

George H. Noell & Joseph C. Witt

Louisiana State University

Jennifer T. Freeland

Indiana State University

Brad A. Dufrene

Mississippi State University

Donna N. Gilbertson

Utah State University

Correspondence concerning this article should be addressed to Gary J. Duhon, School of Applied Health and Educational Psychology, Oklahoma State University, 423 Willard Hall, Stillwater, Oklahoma 74078. E-mail: duhong@okstate.edu

Abstract

This study examined an approach for assessing and intervening with academic concerns that is conceptually derived from the distinction between skill deficits and performance deficits. A class-wide assessment was used to describe the students' functioning in the context of their peers and to help select a target concern for intervention. An individual assessment was also conducted to examine the possibility that goal setting and reward would be sufficient to correct the identified deficit. For two students rewards were effective in the brief assessment, suggesting a performance deficit hypothesis, and for two students they were not, suggesting a skill deficit hypothesis. An extended analysis was then conducted comparing an instructional intervention to one emphasizing motivational variables. In all cases the intervention approach suggested by the initial brief analysis was supported in the extended analysis. Implications for practice and future research are discussed.

Identifying Academic Skill and Performance Deficits:
The Experimental Analysis of Brief Assessments of Academic Skills

Academic concerns are frequently the focus of referrals for consultation services (e.g., Noell, Gansle, & Allison, 1999) and are the most frequent reason referral for evaluation of special education entitlement are made (Lloyd, Kauffman, Landrum, & Roe, 1991). In conducting evaluations to determine entitlement for special education, school psychologists commonly rely on assessment practices that lack demonstrated utility in developing interventions (Witt, Elliott, Daly, Gresham, & Kramer, 1998). Although the development of educational interventions typically occurs in conjunction with the assessment of students' academic functioning, the assessment may not directly inform intervention design. In other words assessment results may identify the need for an intervention, but do not always assist in identification of what that intervention should be. This poor link between assessment and intervention reduces intervention planning to a best guess or trial-and-error procedure, which is inefficient for both students and educators. The inefficiency of trial and error is particularly salient in light of research demonstrating that pretreatment assessments can predict students' differential responding to academic intervention and thus inform intervention design (e.g., Noell, Freeland, Witt, & Gansle, 2001). Developing empirically validated tools (such as pretreatment assessments devices) to guide intervention strategies for academic problems can strengthen school psychologists' ability to efficiently respond to common referrals.

Currently functional or behavior analytic approaches to intervention development have a substantial supportive literature base and provide a direct link between assessment and

intervention. Functional procedures, originally described by Carr and Durand (1985) and Iwata, Dorsey, Slifer, Bauman, and Richman (1982), initially focused on directly testing the impact of environmental manipulations on performance. What resulted was a methodology designed to empirically identify functional or causal relations between behaviors and the events surrounding their occurrence (Martens & Eckert, 2000). Knowledge of this causal relationship makes it possible to implement interventions such as extinction and differential reinforcement procedures which are designed to incorporate the function of existing behavior into the intervention design. However, early functional assessment techniques typically focused on behavioral excesses in individuals with low incidence disabilities (Nelson, Roberts, Mathur, Rutherford, **& Aaroe,** 1999), required highly controlled analog settings, and involved multiple assessment sessions over an extended period of time (Northup et al., 1991) which ultimately limited the applicability of this technology. Subsequently researchers have attempted to broaden the scope of application by evaluating several modifications. As a result current methodological advancements have served to shorten the assessment time (e.g., Derby et al., 1992; Northup et al., 1991), rely less on analog conditions (e.g., Sasso, et al., 1992; Lalli, Browder, Mace, & Brown, 1993; Lerman, & Iwata, 1993), and most recently examine behavioral deficits in subjects with higher incidence concerns (e.g., Daly, Martens, Dool, Hintze, 1998; Eckert, Ardoin, Daly, & Martens, 2002; Noell et al., 1998; Noell et al., 2001). What has emerged from this and similar research are relatively simple, short duration procedures capable of identifying environmental events that can be utilized in the development of interventions designed to increase or decrease the behavior of concern.

Comment [ED1]: This list of authors does not match the reference in the reference list.

Deleted: &

Brief assessments based on experimental analyses of academic concerns can result in

differentiated responding between test conditions (Daly & Martens, 1994; Daly, Martens, Hamler, Dool, & Eckert, 1999; McComas, Wacker, Cooper, Asmus, Richman, & Stoner, 1996). Brief assessments have frequently been conducted within adapted single subject experimental designs. Student performance has typically been assessed under baseline or typical conditions. This performance has then been compared to performance under one or several motivational and/or instructional conditions. Assessments have commonly included a return to baseline and treatment to replicate observed differences between conditions. These assessments or brief experimental analyses provide a direct link between assessment and intervention (Martens, Eckert, Bradley, & Ardoin, 1999).

To guide the selection of treatment components to be used as a part of an experimental analysis, a conceptual and organizational heuristic is necessary. The emerging literature examining brief assessments of academic concerns has drawn from more than one heuristic to guide selection of environmental test conditions (e.g., The Instructional Hierarchy; Haring, Lovitt, Eaton, & Hansen, 1978). The historic distinction between an individual's skills and his or her performance of those skills within psychology may provide a parsimonious and useful organizational scheme to guide at least part of the work of assessing academic concerns. In essence, this distinction acknowledges that individuals who are skilled may not exhibit those skills in some contexts or at some times. Indeed, examination of motivational variables has figured prominently in several studies (e.g., Daly et al., 1998; Eckert et al., 2002; Noell, Roane, VanDerHeyden, Whitmarsh, & Gatti, 2000). Within this framework, the absence of a behavior (e.g., the student fails to write answers to math problems correctly) would indicate either that the learner has not acquired the skill or that he or she has acquired it but simply chooses to do

something else instead (Gresham, 1981; Lentz, 1988; Skinner, 1998). The possibility that poor performance is the result of a skill or performance deficit can be evaluated by assessing the effects of teaching (for skill deficits) or motivational variables (for performance deficits). Skill deficits result from a lack of skill and therefore can be reduced through antecedent manipulations designed to improve the skill level (i.e. instruction). Performance deficits result from a lack of environmental support for performance and can be reduced by providing powerful rewards for academic responding that can compete effectively with available rewards for doing other things (e.g., getting out of doing the math problems). Response to these antecedent or consequence manipulations can then be used as hypotheses about deficit performance. For example, response to an antecedent intervention and not to a reward may result in the hypothesis that any lack of performance in the area assessed is the result of a skill deficit and not a performance deficit. In contrast, response only to the consequence manipulation (i.e. reward) may result in a hypothesized performance deficit. Ideally, these hypotheses could then be used in the development of intervention or as interventions themselves.

Deleted: increase

Eckert, Ardoin, Daisey, and Scarola (2000) and Eckert et al. (2002) evaluated academic responding under antecedent and consequent procedures both in isolation and in combination in an attempt to develop hypotheses about poor reading performance exhibited in elementary subjects. In both studies brief experimental analyses were conducted evaluating the effects of antecedent and consequence procedures on oral reading fluency. Antecedent manipulations such as “repeated readings” and “listening passage preview” were implemented as skill building procedures. Consequences such as “performance feedback” and “contingent reward” were employed in an attempt to impact levels of motivation. Alone and in combination the impact of

these antecedent and consequent manipulations on oral reading fluency were evaluated and hypotheses with regard to the deficit performance were evaluated. Individual differences among students were found in their responsiveness to treatment, supporting the distinction between skill and performance deficits.

Recent research has also demonstrated that differences found during brief assessments comparing differing instructional procedures may be related to responding during an extended intervention. Noell et al. (2001) evaluated skill and performance deficits using brief assessment procedures that consisted of the application of contingent reward (designed to impact motivation) and instructional procedures. These brief assessments (approximately 40 minutes) were implemented across multiple levels of instructional material for students referred due to poor reading performance. Based on student responses to these brief assessments, hypotheses were developed based upon standardized decision rules about the type of deficit (skill or performance) exhibited by the student. These interventions were then tested in an extended analysis (3-4 weeks) designed to evaluate the match between assessment and intervention. Results indicated that the brief assessment matched the results of the extended analysis in 83% of the cases, suggesting the utility of conducting abbreviated assessments to differentiate skill deficits from performance deficits.

Similar to previous work by Noell et al. (2001), this study examines the accuracy of brief assessment in predicting student responses to interventions over extended sessions and time. However, this study extends this line of research in three ways. First, mathematics and writing rather than reading were targeted. Second, much of the assessment and the entire extended analysis were conducted within the students' classroom. Third, interventions tested in the

extended analysis were based on the hypotheses (skill deficit or performance deficit) generated from the results of the initial assessments, but these interventions differed procedurally from the assessment procedures. In other words the interventions implemented as a result of the assessment were of the same class (i.e. instructional or motivational in nature) as the treatments used in the assessment, but were not identical to those used in the brief assessments. Much of the research in this area has evaluated interventions derived from assessment (i.e. Eckert, Ardoin, Daisey, & Scarola, 2000; Eckert, Ardoin, Daly, & Martens, 2002; Noell et al., 2000; Noell et al., 2001) that are procedurally identical to the interventions implemented in a later comprehensive analysis. These early studies were designed to evaluate the degree to which the brief assessment procedure could provide information about generalizability of the intervention over time and stimulus materials. However, the current study sought to determine whether results from a brief assessment were generalizable within a broader class of intervention (i.e. skills intervention or performance intervention) by evaluating the degree to which differentiated patterns of responding to procedurally different skill or performance based interventions corresponded to a priori hypotheses.

A brief in-class group assessment was conducted in the classroom to obtain a point of comparison for additional assessment. This was followed by a brief out-of-class individual assessment that provided an incentive for improved performance to check for the possibility of a performance or motivational problem. Based on the results, a hypothesis reflecting either a performance deficit or skill deficit was advanced. An extended analysis using an alternating treatments design (Barlow & Hersen, 1984) comparing a skill-based intervention with a performance-based intervention was then conducted within the regular classroom utilizing actual

classroom assignments. Results of the brief assessment were then compared to the extended analysis to determine the assessment's ability to predict the students' response to different types of intervention.

Method

Participants and Setting

Participants were four general education students referred by their teacher for poor academic performance. All participants were male, African American, and ranged in age from 8 to 10 years of age. Dallas (8 years, 3 months) was a third grade student referred as a result of failing grades in written language and disruptive behavior. Marty (9 years, 6 months) was a fourth grade student and Jerry (10 years, 1 month) was a fifth grade student. Both Marty and Jerry were referred due to multiple academic concerns, specifically failing grades in math, written language and spelling. Tony (10 years, 7 months) was a fifth grade student referred due to failing grades in mathematics. For Marty and Jerry, who were identified with multiple areas of concern, the teacher selected the specific target activity after reviewing the data from the brief assessment (described below). None of the participants had been previously referred for an evaluation of special education entitlement, and none were receiving any additional services at the time of the study.

The in-class assessment and extended analysis phases were conducted in the general education classroom by the teacher and in the context of the normal classroom routines. The out-of-class assessment was conducted in the school auditorium in the presence of two researchers.

Materials

Materials for this study consisted of reading, math, and writing probes. Reading probes consisted of randomly selected third, fourth, and fifth grade passages from Harcourt Brace Jovanovich Treasury of Literature. The passages were selected employing standard curriculum-based measurement guidelines (see Shapiro, 1989). The mean readability calculated using Flesch-Kincaid readability index for the grade levels included in this study were 3.71 (range 3.11 to 3.94) for third grade, 4.52 (range 4.1-4.96) for fourth grade, and 5.48 (5.1 to 5.97) for fifth grade. Math probes were single skill probes created based on the most recently completed calculation task from previous unit tests identified by the teacher from the Scott Foresman - Addison Wesley Math series. The math skills identified by the teachers consisted of single digit multiplication (3^{rd} grade), 2-digit-by- 2-digit multiplication without regrouping (4^{th} grade) and 3-digit-by-3-digit multiplication with regrouping (5^{th} grade). Probes were created utilizing a Microsoft Excel® spreadsheet specifically configured to generate random numbers for the given problems so that multiple, equivalent probes could be created quickly and easily for repeated use. Writing probes consisted of a sheet of lined paper with a "story starter" written at the top of the page. The story starter was created by the classroom teacher and consisted of an open ended sentence (e.g., "This weekend I would like to …") from which the students would complete a story.

Dependent Variables

The dependent variable for the brief assessment was the amount of academic work completed correctly per minute. For math, the amount of academic work completed correctly was scored for digits correct per minute (dcpm) and derived from a 2 min administration of a grade appropriate probe. A digit was considered correct if the correct number appeared in the

correct column of the answer (Skinner, Turco, Beatty, & Rasavage, 1989). For example, a score of 2 digits correct would be awarded if the answer to "22+9=" was "31". An answer of "21" would result in 1 digit correct because the number in the ones column is correct. An answer of "32" would also result in 1 digit correct because the number in the tens column is correct. Finally, an answer of "22" would be scored as 0 digits correct because neither column contains a correct response. Digits correct per minute were then calculated by dividing the total digits completed correctly by the length in minutes of the math assessment.

For reading and writing, words correct per minute (wcpm) was the dependent measure. A correct word for reading was a word read correctly by the participant within 3 s or self-corrected if mispronounced within 5 sec. A word was scored as an error if the student mispronounced the word, substituted another word, omitted a word, or did not read a word within 3 s. Words correct per minute for reading was calculated by subtracting the number of errors from the total number of words attempted in the 1 min administration. A correct word for writing was a word that was recognizable (i.e., misspelled words were counted as correct as long as they could be correctly identified out of context). For example, the word "friend" was counted correct if spelled as "freind" or "frend" because out of context the misspelling is evident. However, the same word spelled "fernd" would not be counted as correct because it would likely be indistinguishable without context. Words correct per minute were then calculated by subtracting the number of errors from the total number of words written in the 3 minute administration.

The dependent measure for each subject within the extended analysis was the amount of academic work completed correctly within a 10 min work session. During the extended analysis

each student was provided with an academic task specific to the subject area identified as problematic and then given instructions specific to the task. All academic tasks completed in the extended analysis were actual academic tasks assigned by the teacher in which the students were performing poorly.

For Tony, the dependent variable during the extended analysis was based on correct responding to a set of 3-digit-by-3-digit multiplication problems with regrouping. The total number of correct digits during a 10 min work session served as the dependent measure. Digits were scored in the identical fashion as the brief assessment, but the total score reflected digits correct per session (dcps) and not digits correct per minute (dcpm). Digits correct per session were calculated by counting the total number of correct digits written during the session.

The dependent measure for Jerry and Marty during the extended analysis was words written correctly. Jerry and Marty were required to write a story based on a topic provided by the teacher (journal writing). The total number of correct words during a 10 min work session served as the dependent measure. Words were scored in identical fashion to the brief assessment. Upon completion of the 10 min session the writing assignments were scored for words correct per session (wcps) by counting the total number of correct words written within the session.

Finally, for Dallas the extended analysis consisted of detecting grammatical errors in a teacher administered passage and correcting the errors present in the passage within the work session. The percent of total errors detected and corrected (pedc) was the dependent measure. For example, if the passage contained 10 errors and Dallas detected and accurately corrected 8 of the 10 a score of 80% was earned. It is important to note the measurement discrepancy between the brief assessment and the extended analysis for Dallas. Specifically, in the assessment phase

Dallas was asked to complete an open ended writing assignment from a “story starter”, while in the extended analysis he was asked to detect and correct grammatical errors in a given passage. Although these tasks represent different measures, percentage of grammatical errors was chosen for the extended analysis as a result of a teacher request, and because it represented an actual writing task with which Dallas had a history of performing poorly in class. Because the assessment was designed to differentiate between performance and skill deficits in a given academic area for the purpose of intervening with actual academic problems, the teacher identified problem was chosen. The selection of this measure not only adds to the social validity of the dependent measure, but is supported by research that has demonstrated a strong relationship between total words written (assessment measure) and more global criterion measures in written language (Shinn, 1989). A researcher scored all subject work products following each session.

Brief Assessment Procedures

The initial assessment for each student included a class-wide, in-class component and an individual, out-of-class component. The results were used to develop a hypothesis about whether the observed poor performance was due to poor skills (a skill deficit) or poor motivation (a performance deficit).

In-class group assessment. The in-class assessment was conducted to obtain a point of comparison (i.e., baseline) that reflected in-class work in the presence of peers and the student’s teacher under typical conditions. These data would serve as the point of comparison for the subsequent individual, out-of-class assessment and was used to generate the skill deficit versus performance deficit hypothesis (this procedure is described below). The class-wide assessment

was also presented to the teacher to provide a snap shot of the whole class and as an additional aide in selecting a specific target for students exhibiting multiple concerns.

The in-class group assessment consisted of the administration of a 2 min math probe, a 1 min reading probe and a 3 min writing probe by the teacher. The math and writing probes were group administered to the entire class, while the reading probe was individually administered to the target student and two peers in the target student's reading group. The reading probe was not administered to the entire class in order to keep the reading assessment portion as brief as possible. Prior to conducting the class-wide assessment a researcher and the classroom teacher reviewed the procedures for administering the CBM probes. The teacher was then given a script with a set of 15 steps for administering the probes and asked to follow the script while conducting the assessment. The experimenter gathered the materials that corresponded to the grade level assessed and provided sufficient copies to the teacher. The probes were then administered by the classroom teacher using standard CBM administration procedures described in Shinn (1989) to the class as the researcher monitored teacher integrity.

Integrity data for the in-class group assessment were collected for all four in-class group assessments completed. The mean number of correctly implemented steps was 95% (range, 86 to 100%). Due to time constraints the reading probes were only administered by the teacher to the target student and two peers whom the teacher identified as performing in the average range in his or her class.

Out-of-class individual assessment and hypothesis generation. An out-of-class individual assessment was conducted immediately following the in-class group assessment for the referred students to determine if the student's performance during the in-class group assessment would

improve if rewards were introduced. In this session, the target student was given an opportunity to work on a parallel version of each CBM probe completed during the in-class group assessment utilizing the in-class administration procedures with the addition of the reward. Students were given a performance goal for each probe and instructed that if they exceeded this goal they would be allowed to choose a reward from a prize box (i.e. small toy, candy, pencil, etc.). The goal was the score obtained during the in-class group assessment. For example, if the student's score from math during the in-class group assessment was 40 digits correct then his goal for the out-of-class individual assessment became 40.

Based on the results of the out-of-class individual assessment a hypothesis regarding the student's poor performance was generated. This hypothesis was generated by comparing the student's performance from the out-of-class individual assessment to that of performance from the in-class group assessment and to that of instructional placement standards for that grade level. A comparison was made with data from the out-of-class individual assessment to determine if the introduction of rewards improved performance. If the student's performance in the out-of-class individual assessment improved to a marked degree the student's deficit was hypothesized to be the result of a *performance deficit*. An improvement of 50% or more was used as a rule of thumb for identifying marked improvements. In addition, the improvement had to result in performance at or near the instructional placement standards recommended by Fuchs and Deno (1982) for math and reading and Mirkin et al. (1981) for writing. These instructional placement standards represent ranges of performance considered to be essential for progress within grade level. If the reward did not improve performance to this degree the subject's performance was hypothesized to be the result of a *skill deficit*.

Experimental Design and Extended Analysis Procedures

To determine if the outcome of the individual out-of-class assessment suggested the most effective treatment, the relative effectiveness of both treatments was evaluated within an alternating treatments design (Barlow & Hersen, 1984). For example, if the brief assessment suggested a skill deficit, the skill based treatment would be predicted to be more effective than goal setting and reward. The use of ATD allows for comparisons of two treatments or conditions within a single subject without the need for a withdrawal of treatment (Barlow & Hersen, 1984). Conditions were randomized in an alternating fashion for each student. Once a most effective treatment was identified this condition was then implemented for three to four additional sessions to assure that improved performance in that condition was not merely the result of carryover effects.

In the extended analysis, treatments targeting both skill enhancement and motivation were developed and provided for each student. The skill based treatment focused on improving student performance by providing pre-session practice, guided advanced organization, or an instructional prosthetic (see the instruction section below). The performance deficit treatments focused on using goal setting and rewards for achieving those goals (see the reward condition below). All sessions were 10 minutes in length with one session conducted per day for each subject. Each session was integrated within typical classroom activities, consistent with the academic skill addressed. For example, Jerry and Marty completed the writing intervention during their class's journal writing time; Dallas completed his grammatical intervention during the morning grammar exercise for his class; and Tony completed his math intervention during daily math practice activities.

Deleted:

Baseline. The experimenter provided the student his targeted academic task and asked him to complete it. The time allowed to complete the academic task was 10 min; however, this was not communicated to the student. When 10 min elapsed, the academic task was removed and scored by the experimenter. Students were not provided with any feedback on their performance by the experimenter.

Performance Feedback. This condition was the same as the baseline condition; however, the experimenter scored the academic task immediately and the student was told his score. This condition was introduced to evaluate whether feedback alone was sufficient to improve performance. Isolating the effects of feedback was important because grading and feedback are part of both the reward and instruction procedures, but were not the focus of this study. In essence this served as an additional control condition, designed to ensure that the treatment conditions were not confounded by the application of feedback alone.

Reward. Before the academic task was presented, a goal was provided to the student who was told that if he exceeded the goal, he would be allowed to select a reward from the prize box. The goal for each student was based on either the median of the three previous scores from the reward condition or a combination of reward and baseline if less than 3 reward sessions had been conducted. This goal setting procedure is known as percentile shaping (Galbicka, 1994). Following the reward sessions, the experimenter collected the academic task, scored it, and told the student his score. If the student exceeded the goal, he was immediately allowed to choose a reward from the prize box. If the student did not meet his goal, he did not receive a reward.

Instruction. Prior to the academic task session, each student met with an experimenter outside of the classroom. The student was provided with pre-session practice, an advanced

organizer, or an instructional aid. These interventions were designed to provide the student assistance at the time of the session to improve responding during classroom instruction or provide compensatory support for weak skills. For Marty and Jerry, whose target task was writing production, the experimenter worked with the students prior to the session to help them plan for their writing and complete an outline (Kellogg, 1988). The experimenter gave the student the story starter (e.g., "Today at recess I....") and then helped the student generate a list of main ideas for use in response to the story starter. The outline was then available for reference during the writing session.

For Tony, whose task was to multiply three-digit-by-three-digit numbers, the experimenter provided a table of single digit multiplication facts and demonstrated problem completion using the multiplication facts. This intervention was chosen because after reviewing the results of the assessment it was discovered that Tony did not know his multiplication facts, but did understand the process required for solving the three-digit-by-three-digit multiplication problems. The fact sheet was designed to act as a bridge to remediate Tony's math performance in the instructional condition in a way that could easily be removed in the reward condition, reducing the likelihood of carryover effects affecting the results. Upon completion of the study Tony was taught his multiplication facts by an experimenter utilizing a time delayed flashcard procedure (Fleischner & Manheimer, 1997).

Finally, for Dallas, the task was a proof reading task in which he was asked to detect grammatical errors in sentences and correct them. Prior to each session the experimenter reviewed the previous day's assignment identifying the errors with Dallas and correcting them (Espin & Sindelar, 1988). The students then returned to the classroom and the same procedures

described in Performance Feedback condition were followed.

Interobserver Agreement

Inter-scorer agreement was calculated for the academic work completed correctly within a 10 min session. A second researcher recalculated the scores for 35% of the sessions completed for each participant. Inter-scorer agreement was calculated by dividing the total number of agreements in (a) number of digits correct, (b) words correct, or (c) errors detected and corrected for each academic task by the number of agreements plus disagreements, and then multiplying the product by 100. Average inter-scorer agreement was 98% (range, 83 to 100%) for academic work completed correctly within a 10 min session across all four subjects.

Results

Results of the brief assessments (both the in and out of class portions) for all participants are reported in Table 1. Percent improvement and instructional placement standard as well as the hypothesized deficits are also reported. Results of the extended analyses are displayed in Figure 1.

Marty wrote 11 words per minute during the in-class assessment and with the addition of reward during the out-of-class assessment he wrote 12 words per minute. Performance in both settings was well below the instructional range and reward produced only an improvement of only 9%, which was interpreted as suggesting a skill deficit. Marty' performance during the extended analysis is presented in the first panel of Figure 1. During the initial baseline of the extended analysis Marty's performance was stable (M = 33.0 words correct per session). Performance feedback was associated with a decrease in performance (M = 20.3 wcps). During the extended analysis the instruction condition (M = 76.5 wcps) produced superior results to the

reward condition (*M* = 51.3 wcps). These differences produced effect sizes of 4.73 for instruction over baseline and 2.05 for instruction over reward. The results of the extended analysis indicate that an instructional intervention was more effective for Marty, which is consistent with the findings of the brief assessment.

Tony completed 8.5 digits correct per min during the in-class group assessment and was well below the instructional range for math. During the out-of-class individual assessment with reward he completed 4 dcpm. This reflected a 53% deterioration in performance and was interpreted as suggesting a skill deficit. During initial performance feedback phase of the extended analysis (see panel 2 Figure 1) Tony's performance exhibited a downward trend. His mean performance in performance feedback was 71 digits correct per session (dcps). Tony was not exposed to the baseline only condition as the result of an error made by one of the researchers in providing performance feedback during the initial sessions. As a result it is unknown what Tony's baseline performance would have been without performance feedback. However, because Performance Feedback had little or no impact above baseline for the three remaining subjects, it is reasonable to assume that Performance Feedback had little impact on Tony's performance as well.

During the extended analysis for Tony, initial sessions were no different than baseline. However, after exposure to both conditions the instruction condition (M = 116.1 dcps) produced superior results to the baseline mean and the reward condition (*M* = 90.6 dcps). These differences produced effect sizes of 3.35 for instruction over baseline and 1.56 for instruction over reward. Although there was initial overlap in the data paths, it may be that Tony was not discriminating the conditions very well. As the analysis continued, the trend in the data thereafter

indicates that the instructional intervention was more effective for Tony, which is consistent with the findings of the brief assessment.

Jerry wrote 22 words per minute during the in-class group assessment, which is considered frustrational for 5th grade writing. During the out-of-class individual assessment with reward he wrote 43 words per minute. The difference between the in-class group assessment and the out-of-class individual assessment reflects a 95% improvement in performance and is very close to the lower limit of the instructional range of 49 words per minute and therefore was interpreted as a performance deficit.

Jerry's performance during the extended analysis is presented in the third panel of Figure 1. During initial baseline of the extended analysis Jerry's performance exhibited a downward trend with a mean of 21 words correct per session (wcps). Performance feedback was associated with a modest increase in performance (M = 27.5 wcps). During the extended analysis, the reward condition (M = 140.3 wcps) produced superior results to the instruction condition (M = 114.8 wcps). These differences produced effect sizes of 7.12 for the reward condition over baseline and 1.85 for reward over instruction. Like that of Tony the initial data paths overlap, but after a few sessions the paths separate. In Jerry's case reward and instruction initially produce similar results, however performance in the instructional condition remains flat, while the reward condition reflects an upward trend, which produced a separation in the data paths. It is noteworthy that the improvement in the reward condition occurred on the first session, prior to any instruction sessions, and that it persisted for the four sessions following the cessation of instruction sessions. The results of the extended analysis indicate that a motivational intervention was more effective for Jerry, which is consistent with the findings of

the brief assessment.

During the in-class assessment Dallas performed within the frustrational range with 17 words correct per minute. During the out-of-class assessment with reward he wrote 32 wcpm. This reflected an 88% improvement in performance, and, like Jerry, was considered very close to the instructional range of 34 wcpm for his grade level. Based on these results a hypothesis of a performance deficit was forwarded.

Dallas's performance during the extended analysis is presented in the final panel of Figure 1. When given a passage with errors to detect and correct, Dallas corrected a similar percentage of errors across the initial baseline (*M* = 45.3 percent errors detected and corrected) and performance feedback (*M* = 40.5 pedc) conditions. During the extended analysis the reward condition (M = 66.0 pedc) produced superior results to the instruction condition (*M* = 46.2 pedc). These differences produced effect sizes of 1.39 for reward over baseline and 1.89 for reward over instruction. It is interesting to note that Dallas's performance during the instructional condition was similar to his performance during baseline. Aside from one overlapping data point, the results of the extended analysis indicate that the motivational intervention was successful in improving performance for Dallas over baseline levels while the instructional intervention was not. These findings are consistent with the findings of the brief assessment.

Discussion

This study examined the extent to which the hypotheses suggested by a brief, relatively simple assessment would predict students' response to a functionally relevant instructional or motivational intervention. The hypotheses and interventions were developed based on the heuristic distinction of skill versus performance deficits. The results for all four participants

were very encouraging. This study replicates previous research by examining brief assessments of students' instructional needs. Similar to previous research (Daly & Martens, 1994; Daly, et al, 1999; Eckert et al., 2002), the results of this study suggest the potential utility of brief assessments using planned test conditions to guide selection of appropriate intervention elements.

This study extends previous research in several ways. First, part of the brief assessment and all of the extended analyses were conducted in the students' classroom by the students' classroom teachers. Previous research in this area has typically relied on out-of-class assessments and interventions (e.g. Daly, et al, 1998; McComas et al., 1996). Second, this study targeted mathematics and written language performance. Previous research in this area has targeted reading. Third, this study provides promising evidence that hypotheses derived from brief assessments regarding skill-based or performance-based intervention classes may be generalizable to some procedural variations within those classes. This finding is strengthened by the consistency of results across academic areas—writing and math. Previous research examining congruence between the brief and extended assessments has used the same procedures across the brief and extended assessments (Noell et al., 2001). In this study the goal setting and reward procedures differed procedurally but not functionally between the brief and extended assessments. Finally, it sheds further light on how teachers might be able to simplify their task in some situations based on the results of brief assessments. The use of motivational variables as part of the brief assessment may prevent the unnecessary application of antecedent instructional interventions, which presumably involve more effort on the part of the teacher. In other words, a motivational intervention may be sufficient to maintain performance and additional instruction

may not be needed. This conclusion can be reached quickly and simply.

This study suggests that observations of student performance under controlled conditions in which a contingency is in place for the desired behavior can produce valuable assessment data. This information can strengthen the problem solving, problem analysis or intervention design process by helping practitioners to focus efforts on appropriate levels of tasks and interventions with a potentially higher probability for success (Daly et al., 1999; Eckert et al., 2002; Noell et al. 2001). An additional advantage for practice is that the assessments are relatively brief and simple. This type of assessment can readily be integrated into a variety of behavioral and curriculum based assessment approaches. The assessment procedures described herein could be integrated into more broad-spectrum approaches such as Behavioral Consultation (Bergan & Kratochwill, 1990) or more specific approaches such as Curriculum Based Assessment (Shapiro, 1996) or Curriculum Based Evaluation (Howell & Nolet, 2000). The assessment of motivational variables should not be conceptualized as an activity that would compete with existing practices. It should be an assessment activity that practitioners can integrate into their ongoing assessment and intervention activities.

Although the results are encouraging several limitations and areas for future research should be noted. While the distinction between skill and performance deficits has been recognized within psychology and its application to school psychology has been discussed conceptually for some time, research examining the utility of the distinction is very limited. Important factors such as the accuracy of assessment, predictive validity and treatment utility of assessments based on this distinction, with a few exceptions, have been primarily speculative rather than evidence based. Therefore the methods by which it is examined have not been

standardized to any great extent. The approach taken within this study is only one example that may have the potential for distinguishing between performance and skill deficits. One advantage of the approach taken here is the efficiency with which these procedures can be conducted. With very little training and support it is likely that teachers could administer similar types of performance and skill assessment. Second, results of the brief analysis can be used to develop treatments that have an evidential basis for use.

The practicality of the assessment procedures used has four primary limitations. First, a skill deficit was inferred for Marty and Tony by default. In other words when Marty and Tony did not respond to the reward with improved performance they were classified as having a skill deficit. It is possible that for Marty and Tony the reward offered was not reinforcing. If this were the case, performance would not improve with the application of reward and the classification of skill deficit would be based on the faulty assumption that it was the class of intervention that was not functional while in reality it was the quality of the reward that was responsible for the lack of performance. The elimination of default categorizations as well as a more exhaustive evaluation of both instructional and motivational manipulations could serve to increase the accuracy with which reasons for academic performance deficits exist and, more importantly, facilitate the choice of effective interventions. Second, the default diagnosis of skill deficit does not reveal which types of instruction are appropriate. If a skill deficit is identified, the results of the assessment provide no information about which instructional procedure may be effective for the identified deficit. However, future attempts could incorporate an instructional heuristic in order to inform instruction. Similar to work by Daly & Martens (1994) and Daly et al. (1998), an instructional heuristic such as the Instructional Hierarchy (originally described by

Haring, Lovitt, Eaton, and Hansen, 1978) could be incorporated into this type of brief analysis to add specific recommendations for students identified with instructional needs. Other researchers (e.g. Eckert, et al., 2000, Bonfiglio, Daly, Martens, Lin, & Corsaut, in press) have incorporated instructional as well as performance based components in isolation and in combination to evaluate student responses to reading intervention. These methods may also provide a better understanding of student performance as well as more precise recommendations for treatment. This increase in precision, however, comes with a cost in terms of additional time, effort and training necessary to complete the assessment. Third, a potential confound exists as a result of the shift in assessment methodology within the initial assessment condition. The out-of-class assessment resulted in a change in multiple aspects of the environment (both a location change and the addition of a reward component). This shift represents a lack of control that may affect participant performance in ways that are not solely attributable to the reward condition. Although results of the study make this less likely it should still be noted as a potential confound.

A fourth limitation is the lack of information with regard to the combined effect of the motivational and instructional variables. The possibility exists that some academic deficits are the result of a combination of performance and skill deficits. The current assessment is not designed to address this possibility and enhancements should be made to provide a more detailed examination of the treatments in combination and alone. Finally, it is possible that the results would not generalize across other procedural variations of interventions. The extended analysis does not currently test other possible interventions to determine their potential impact. However, the design of the extended analysis could be modified to evaluate the generalizability of treatments within a conceptual class of interventions. For example, if through the extended

analysis the instructional modifications are identified as an effective treatment method, other similar instructional modifications could be evaluated to determine if they too can produce similar improvements in behavior.

A limitation of this study not related to the assessment procedures was the lack of treatment integrity data collected during the extended analysis. This lack of integrity data makes it impossible to know if interventions were conducted exactly as planned. Another possible limitation was the use of the percent of total errors detected and corrected for Dallas as a dependent measure. This is a unique conceptualization of grammatical accuracy and was chosen mainly for its match with the original teacher concern. However, it represents a shift in measurement methods for Dallas across the brief and extended analysis. Another limitation is the extent to which the results in the alternating treatments design may have been confounded by carryover effects. It is possible that the application of an intervention in one condition influenced performance in the other condition. Although the interventions were procedurally distinct and the application of the treatments was counterbalanced (Barlow & Hernsen, 1984), the potential carryover effects cannot be completely dismissed. The small size of the participant sample and the academic tasks should also be considered a limitation. The procedures presented here should be tested across a broader range of students and academic tasks.

As mentioned earlier, brief analyses of academic deficits are not intended to be utilized in isolation, but as part of a more comprehensive assessment model. However, the information obtained from methods such as those described in this study can provide important insight into student performance and lead directly to changes in instructional programming with little inference on the part of the practitioner. Future application of the brief analysis of academic

deficits could serve to provide the initial direction for which practitioners pursue more detailed and thorough analysis.

References

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon Press.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy.* New York: Plenum Press.

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional management decisions. *Exceptional Children, 42,* 233-238.

Bonfiglio, C. M., Daly, E.J., III, Martens, B. K., Lin, L. R., & Corsaut, S. (in press). An experimental analysis of reading interventions: Generalization across instructional strategies, passages, and time. *Journal of Applied Behavior Analysis.*

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18,* 111-126.

Daly, E. J. III, Lentz, F. E. Jr., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11,*, 369-386.

Daly, E. J., Martens, B. K., Hamler, K. R., Dool, E. J., Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94.

Daly, E. J., Martens, B. K., Dool, E. J., Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 203-218.

Daly, E. J., Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior*

Analysis, 27, 459-469.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Derby, K. M., Wacker, D. P., Sasso, G., Steege, M., Nothup, J., Cigrand, K., & Asmus, J. (1992). Brief functional assessment techniques to evaluate aberrant behavior in an outpatient setting: A summary of 79 cases. *Journal of Applied Behavior Analysis, 25,* 713-721.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: The use of brief experimental analysis and single-case designs. *Psychology in the Schools, 37,* 463-474.

Eckert, T. L., Ardoin, S. P., Daly, E. J., III, & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35,* 271-281.

Espin, C. A., Sindelar, P. T. (1988). Auditory feedback and writing: Learning disabled and nondisabled students. *Exceptional Children, 55,* 45-51.

Fleischner, J. E., Manheimer, M. A. (1997). Math interventions for students with learning disabilities: Myths and realities. *School Psychology Review, 26,* 397-414.

Fuchs, L.S. & Deno, S.L. (1982). *Developing goals and objectives for educational programs,* Washington, DC: American Association of Colleges for Teacher Education.

Galbicka, G. (1994). Shaping in the 21st centruy: Moving percentile schedules into applied settings. *Journal of Applied Behavior Analysis, 27,* 739-760.

Gickling, E., & Havertape, J. (1981). *Curriculum- based assessment.* Minneapolis, MN: National School Psychology Inservice Training Network.

Gickling, E. E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 587-595). Washington, DC: National Association of School Psychologists.

Gresham, F.M. (1981). Assessment of children's social skills. *Journal of School Psychology 19,* 120-133.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom.* Columbus, OH: Merrill.

Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd ed.). Belmont, CA: Wadsworth.

Iwata, B., Dorsey, M., Slifer, K., Bauman, K., & Richman, G. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities, 2,* 3-20.

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 355-365.

Lalli, J. S., Browder, D. M., Mace, F. C., & Brown, D. K. (1993). Teacher use of descriptive analysis data to implement interventions to decrease students' problem behavior. *Journal of Applied Behavior Analysis, 26,* 227-238.

Lentz, F. E. (1988) Effective reading interventions in the regular classroom. In J. L. Graden, J. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 351-370). Washington, DC: The National Association of School Psychologists

Lerman, D. C., & Iwata, B. A. (1993). Descriptive and experimental analyses of variables maintaining self-injurious behavior. *Journal of Applied Behavior Analysis, 26,* 293-319.

Lloyd, J. W., Kauffman, J. M., Landrum, T. J., & Roe, D. L. (1991). Why do teachers refer pupils for special education? An analysis of referral records. *Exceptionality, 2,* 115-126.

Martens, B. K., Eckert, T. L. (2000). The essential role of data in psychological theory. *Journal of School Psychology, 38,* 369-376.

Martens, B. K., Eckert, T. L., Bradley, T. A., & Ardoin, S. P. (1999). Identifying effective treatments from a brief experimental analysis: Using single-case design elements to aid decision making. *School Psychology Quarterly, 14,* 163-181.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996). Brief experimental analysis of stimulus prompts for accurate responding on academic tasks in an outpatient clinic. *Journal of Applied Behavior Analysis,* 29, 397-401.

Mirkin, P.K., Deno, S. L., Fuchs, S. L., Wesson, C., Tidal, G., Marson, D., & Kuehnle, K. (1981). *Procedures to develop and monitor progress on IEP goals.* Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Nelson, J. R., Roberts, M. L., Rutherford, R. B., Mathur, S. R., & Aaroe, L. A. (1999). A statewide survey of special education administrators and school psychologists regarding functional behavioral assessment. *Education & Treatment of Children, 22,* 267-279.

Noell, G. H., Gansle, K. A., & Allison, R. (1999). Do you see what I see? Teachers' and school psychologists' evaluations of naturally occurring consultation cases. *Journal of Educational and Psychological Consultation, 10,* 107-128.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H.,

Gilbertson, D. N. & Northup, John. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659-663.

Noell, G. H. Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39,* 335-355.

Noell, G.H., Roane, H.S., VanDerHeyden, A.M., Whitmarsh, E.L. and Gatti, S.L. (2000). Programming for communication in the classroom following an assessment of skill and performance deficits. *School Psychology Review, 29,* 429-442.

Northup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis, 24,* 509-522.

Sasso, G. M., Reimers, T. M., Cooper, L. J., Wacker, D., Berg, W., Steege, M., Kelly, L., & Allaire, A. (1992). Use of descriptive and experimental analysis to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis, 25,* 809-821.

Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and interventions.* (2nd Ed.) New York: Guilford.

Shinn, M.R. (1989). *Curriculum-based measurement: Assessing special children,* New York: Guilford Press.

Skinner, C. H. (1998). Preventing academic skill deficits. In T. S. Watson & F. M. Gresham (Eds.) *Handbook of child behavior therapy* (pp. 61-82). New York: Plenum Press.

Skinner, C. H., Turco, T. L., Beatty, K. L., & Rasavage, C. (1989). Cover, copy and compare: A method for increasing multiplication performance. *School Psychology Review, 18,* 412-420.

Witt, J.C., Elliott, S.N., Daly III, E.J., Gresham, F.M., & Kramer, J.J. (1998). *Assessment of at-risk and special needs children* (2nd ed.). New York: McGraw-Hill Higher Education.

Table 1.

Assessment results for targeted skill from both in and out-of-class assessment including percent of improvement, instructional placement standard and hypothesized deficits.

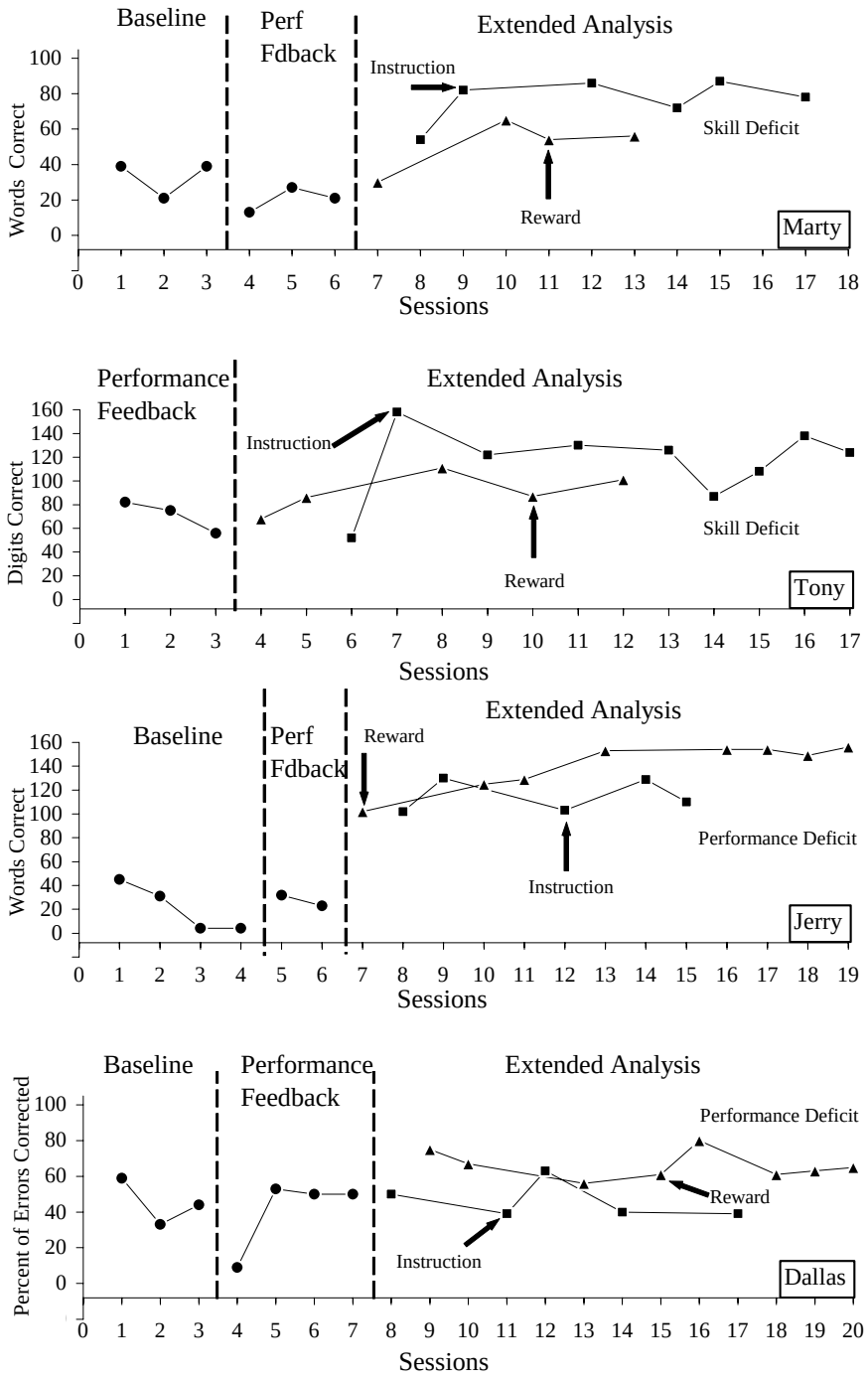
Name	Grade	Referral Subject	In-class score	Class median	Out-of-class score	Percent Increase	Instructional range [a]	Hypothesized Deficit
Marty	4th	Writing	11 wcpm	15 wcpm	12 wcpm	9%	41+	Skill Deficit
Tony	5th	Math	8.5 dcpm	17.5 dcpm	4 dcpm	-53%	20–39	Skill Deficit
Jerry	5th	Writing	22 wcpm	14 wcpm	43 wcpm	95%	49+	Performance Deficit
Dallas	3rd	Writing	17 wcpm	19 wcpm	32 wcpm	88%	34+	Performance Deficit

[a] Instructional placement standards were derived from Fuchs and Deno (1982) for math and Mirkin et al. (1981) for writing.

Figure Caption

Figure 1. Academic performance for each subject in the extended validation.

Figure 1


Baseline
Perf Fdback
Extended Analysis
Instruction
Skill Deficit
Reward
Marty
Words Correct
Sessions
Performance Feedback
Extended Analysis
Instruction
Skill Deficit
Reward
Tony
Digits Correct
Sessions
Baseline
Perf Fdback
Extended Analysis
Reward
Performance Deficit
Instruction
Jerry
Words Correct
Sessions
Baseline
Performance Feedback
Extended Analysis
Performance Deficit
Reward
Instruction
Dallas
Percent of Errors Corrected
Sessions