Using Experimental Analysis to Identify Reading Fluency Interventions: Connecting the Dots

Edward J. Daly III, Melissa Andersen, Valerie Gortmaker and April Turner
University of Nebraska-Lincoln

Abstract

This article reviews the conceptualization, measurement, and design of brief experimental analysis for oral reading fluency problems. It presents examples from the literature of how brief experimental analysis results have been used to generate effective treatments for a variety of different applications (e.g., parent tutoring, small group, self-managed interventions). It also describes three different approaches investigators have taken to conducting brief experimental analyses. Finally, the article describes a method for conducting a single trial brief experimental analysis that will allow practitioners to quickly and efficiently identify potential interventions designed to address skill and performance based oral reading fluency deficits. Limitations and areas where future research is needed are discussed.
Keywords: Academic performance, curriculum-based measurement, experimental analysis, oral reading fluency, stimulus control.

Behavior Analysts frequently have legitimate reason to bemoan how their work is characterized by others. Their contributions are often marginalized and criticized in mainstream educational circles by individuals who may prioritize paradigm preferences and philosophical biases over the quality of results produced by differing educational methods. Yet, from time to time a remark from outside of behavior analytic circles rings true and, if we are not too quick to dismiss it, may provide us with fresh insight into the nature of our work. In this case, the comment came several years ago from a then 4-year old daughter of a doctoral student in school psychology. When she saw her mother's graphs of data, the daughter exclaimed, "Oh! It's connect-the-dots!" (Christine Bonfiglio, personal communication, 2002). Lest this commentary be dismissed as merely a cute reflection of an innocent who knew nothing about behavior analytic practices, we wish to point out that this little girl's point reveals a profound truth about our work. Her understanding may be greater than we are willing to give her credit for, even if she understood nothing about principles of reinforcement or stimulus control.

To this little girl, the activity of connecting the dots was sure to produce a picture out of an otherwise incomprehensible jumble of markings on the page. To the behavior analyst, the markings (dots) represent snapshots of behavior at various points in time and under various conditions. And just as the little girl confidently assumed that someone created an order to the dots for her to discover if she persisted with the task, behavior analysts confidently assume that there are predictable functional relationships that will allow them to put meaning and order to the picture in spite of the myriad of variables that may be operating to distract or overwhelm their attention. It is our intention in this paper to provide guidance in how to bring order to the dots associated with oral reading fluency problems. When analyses are appropriately structured, the "connections" between the dots provide valuable stimuli that can be used to occasion more effective teaching methods.

The remainder of this paper will be devoted to unfolding more completely what exactly it is that we are or should be assessing for reading fluency problems and how to fit direct measures of student reading performance into experimental analyses that can inform intervention selection in classrooms and schools. To this end, after conceptualizing the task, we review the literature on experimental analyses whose chief purpose has been to facilitate intervention selection (as opposed to a broader or more comprehensive review of experimental analyses of academic performance). Finally, we outline some ways in which these methods can be used efficiently by educational personnel to resolve reading fluency problems.

Word Reading and Its Measurement

In order for the dots to represent a meaningful educational picture, we must first ask what the conditions are that generate the dots in the first place. The dots are the product of a measured interaction between the student and pre-planned environmental stimuli, which would be an academic task of some type in the case of academic performance. The task and structured assessment conditions rely on an appropriate conceptualization of what the dots are supposed to mean in order to be interpretable. The task itself which is chosen as the source of stimulus materials during assessment obviously has a significant influence on the value of the dot in the overall picture. It should contain the most critical features of the curriculum if the results are to reflect an educationally relevant outcome. For instance, the assessor may choose to assess reading in fourth grade reading materials because the student is in fourth grade. There is more to working out the conceptualization of the dot, however. To complete the analysis, we must turn to the response.

For academic skills, each dot represents the degree to which the task exerts stimulus control over the appropriate response. For example, textual stimuli occasion a reading response of some type. If we ask a student to read aloud when we present him or her with a text, we expect to hear words that correspond exactly to the printed stimuli. For each word, there is one and only one response that is correct, and, therefore, the text should always occasion the same response. The measurement then is an indication of the presence or absence of stimulus control. However, responses within a response class vary in a number of ways across opportunities. The assessment conditions also are presumed to capture some quantifiable *dimension* of responding that is important and which is expected to change over time if response strength is initially weak or even inexistent. For instance, a measurement system in reading might reflect the number of correct responses (frequency) or it might reflect the speed of correct responses (rate or fluency). With effective teaching, response strength should increase over time and across different dimensions. A useful measure will accurately indicate the degree to which this is occurring, giving meaning to the dots on the page.

Measuring oral reading fluency. Fluency is a particularly useful dimension of behavior to measure. Reflecting a combination of accuracy and speed, fluency has proven to be a valid and sensitive indicator of instructional outcomes (Binder, 1996). Indeed, because of its critical role in reading acquisition, *oral reading fluency* has been established as a legitimate instructional target in its own right (National Reading Panel, 2000; Snow, Burns, & Griffin, 1998). Research supports the relationship between reading fluency and overall reading ability, including comprehension (Cunningham & Stanovich, 1998; Meyer & Felton, 1999). Oral reading fluency is a prerequisite to independent comprehension. When children laboriously decipher words in text, their decoding competes with comprehension efforts and impairs their ability later to give a verbal report of what they read.

Oral reading fluency has been operationalized into standardized procedures for creating interpretable dots, referred to in the literature as curriculum-based measurement (CBM; Shinn, 1989). Scored as correctly read words per minute (CRW per min), CBM involves repeated measurement of student proficiency in basic academic skills over time using standardized directions and brief fluency timings (Hintze, Daly, & Shapiro, 1998). CBM was developed as a general outcome measure and provides a reliable, valid, sensitive, and efficient procedure for obtaining performance data that may be used to evaluate instruction (Fuchs & Fuchs, 1999). CBM has a wide variety of applications. For instance, it can be used to model growth longitudinally (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), develop and maintain appropriate student goals (Fuchs, Fuchs, & Hamlett, 1989; Fuchs & Shinn, 1989), and provide information about how to modify instruction (Deno, Fuchs, Marston, & Jongho, 2001; Fuchs, Fuchs, Hamlett, & Allinder, 1991).

Measuring generalization of word reading. Measuring response fluency alone is not sufficient for producing meaningful assessment results. An even more important question is the generality of the behavior (Stokes & Baer, 1977). If a student is able to reliably and quickly read a word in a single text, but cannot read the word when it appears in other texts, the response is of limited utility to the student. If the generality of word reading is not measured, then it is unlikely that educators will take steps to program for it (Stokes & Baer, 1977). Generalization of word reading can be seen as occurring in two forms. In the first case, a student who learns to read a word in one text may then be able to read it in other texts. Change in word order across texts serves as perhaps the most critical change in stimulus conditions that allows us to conclude that generalization occurred. (Although, maintenance of the response across time is another important dimension of behavior as well and is confounded with changes in word order in the example.) Generalization can also occur when a student reads a word he or she never read before in part as a function of having learned to read other words. For instance, if a student learns to read a phonetically regular word (e.g., "box") and responds correctly in the presence of an untrained word (e.g., "mop"), he or she is said to have generalized. This same form of generalization also can be seen in words that do not share stimulus properties, as phonetically similar words do. Surely, teachers do not teach all possible stimulus-response relationships, but there are plenty of students who somehow come to learn them (Alessi, 1987). For example, a second grade student may show a generalized increase in word reading fluency across curriculum items before the teacher even teaches many of the words.

Oral reading fluency assessments can provide a more complete account of behavior if they systematically address the degree of generalization being sampled as dots are generated. According to Alessi (1987), assessment of instructed stimulus-response relationships provides information about student mastery and assessment of uninstructed stimulus-response relationships provides information about generalization. In other words, when responding is measured in directly taught material, the results indicate mastery of what was taught. When responding is measured either for untaught but functionally equivalent responses or for taught responses under different stimulus conditions, the results indicate generalization of responding. Fuchs and Deno (1991) refer to the former practice as *specific subskill mastery measurement* and to the latter as *general outcome measurement*. General outcome measurement is the stronger measurement model for the generalized outcomes that teachers and other educators desire for students, and should therefore serve as the ultimate criterion of instructional effectiveness.

General outcome measurement with untaught stimulus-response relationships (e.g., many untaught words from the curriculum) will be the hardest level of generalization to achieve with students, especially those referred for reading problems. Improvements are likely to show up on graphs more slowly, if at all. Yet, there may be other important types of generalization occurring, and general outcome measurement may actually underestimate a student's responsiveness to instruction. One way in which reading fluency assessments have been structured to provide an index of generalization is to manipulate word overlap between passages used for instruction and those used to assess the effects of instruction (Daly, Martens, Kilmer, & Massie, 1996). Word overlap refers to the amount of word commonality across texts (expressed as a percentage of the same words that appear in both an instructional text and a text used to assess instruction). Passages are created in which many of the same words appear, but which are written as different stories. Stimulus conditions are varied (i.e., sequences of words and therefore meaning) while actual words appearing in both texts remain highly similar. Therefore, assuring high word overlap (e.g., greater amount of identical words between instructional and assessment passages) is one method for estimating the ability of instruction to produce generalized increases.

Use of high word overlap passages is perhaps an intermediate form of generalization which may be more sensitive to instructional effects than traditional general outcome measurement practices like formative evaluation in non-curricular materials that have low word overlap with what is taught. Measuring generalized performance increases within experimental analyses may improve the ability of experimental analyses to identify potentially useful interventions that can be applied in natural settings.

Without careful planning for evaluating generalization of effects, successful treatment of academic difficulties may fail to be relevant to the needs of children experiencing academic difficulties.

Establishing Functional Relationships for Oral Reading Fluency

Our inquiry thus far has been into what the dot means from a behavior analytic perspective. But, we have looked at only part of the picture. When we inquire about the conditions that yield dots, we must also ask what happens *between* the dots to fully interpret their meaning. Between each dot, there will be variations in the stimuli (even if only as a function of time and previous exposure to the stimuli). When the dots increase steadily but the conditions of assessment do not change functionally across sessions, we infer that stimulus control is growing stronger. Joe Witt once said, "The goal of academic intervention is to get the dots to go to the top of the page" (Joseph Witt, personal communication, 1997). The connections between dots inform our interpretation of functional relationships. There are some situations in which we might want to purposefully exploit the direction of the dots by introducing variations in stimulus conditions between sessions. These intentional variations in stimulus conditions should influence the direction of the dots both up and down across assessment sessions. Planned variations in stimuli lie at the heart of the essence of experimental analysis. Variables are directly manipulated to determine the degree of stimulus control achieved across conditions. The results are used as a basis for deciding how to change instruction in the classroom, bringing a picture to the forefront that can help educators prioritize instructional variables for subsequent classroom intervention. In other words, experimental analysis contributes to the overall goal of making the dots reach the top of the page by directly controlling when they go up or down (as a function of instructional conditions and control conditions), allowing the clinician to draw an individualized picture for each case.

In any discussion of establishing functional relationships through experimental analysis, it is essential to relate all of the variables back to the natural environment, which is where those relationships must gain a foothold in order for the student to be successful in the curriculum. With this in mind, we point out that the teacher should have at least three objectives for the classroom for academic subjects like reading, all of which influence the teacher to change academic stimuli in very important ways so that stimulus generalization can be achieved. In promoting student learning, the teacher first aims to reduce and eventually eliminate response prompts necessary to help students make correct responses. Unless this is done, student responding will never fully come under the control of the target academic stimuli. Stimulus control is a prerequisite to stimulus generalization (Shahan & Chase, 2002). Second, the teacher progressively increases task difficulty level and complexity to meet the objectives of the sequential curriculum, which requires stimulus generalization for repertoires taught earlier in the curriculum. Finally (and related to the second objective), the teacher programs instructional activities to progressively approximate "real world" applications outlined as the outcomes of the curriculum. To be successful students must persist in responding correctly (i.e., according to the response demands and critical features of the academic stimuli) to all these stimulus changes. Ultimately, teachers are preparing students to display sophisticated behavioral repertoires in future environments (e.g., college, work settings, personal lives) for which the contingencies may not be altogether clear for any given student. In this process, careful attention is given to the development of stimulus control and then stimulus generalization (so that students' behavioral repertoires are robust in the post-education environment), and any experimental analyses of student performance should be aligned accordingly if the results are to be generalizable to the classroom.

Skill versus performance deficits. It is when a student fails to get the right answer in spite of instructional efforts that an experimental analysis may be called for. An experimental analysis is carried out to establish the variables that will bring about stimulus control and stimulus generalization. Stimulus control itself comes about through differential reinforcement (Catania, 1998). Therefore, when there is a student problem the contingencies are not supporting the occurrence of the desired academic response.

There are two probable reasons why this is happening. In the first case, the consequences for responding are not effective reinforcers for the appearance of the behavior (even though the response repertoire exists). In other words, the student would give a correct response if stronger or more desirable consequences for behavior existed. In the second case, the consequences may be potentially effective (strong enough or desirable enough), but the response repertoire is not of sufficient strength for it to appear in the presence of those consequences. In this situation, antecedent stimuli that are naturally present do not serve as effective prompts for the response either. The former scenario is a performance deficit and the latter scenario is a skill deficit (Lentz, 1988).

Identifying the type of problem is helpful for resolving it. For instance, manipulation of consequences should be sufficient to resolve a performance deficit. In the case of a skill deficit, additional antecedent stimuli in the form of response prompts will be necessary to occasion responding in the presence of natural stimuli (e.g., words in a text) so that it can be reinforced. Corrective feedback will also play a critical role in the formation of appropriate discriminations as well. The types of instructional response prompts and consequences necessary can be differentiated according to a heuristic referred to as the Instructional Hierarchy (IH; Daly, Lentz, & Boyer, 1996; Haring, Lovitt, Eaton, & Hansen, 1978). The IH guides how to increase response frequency for behavioral deficits. Modeling and error correction (involving consequent modeling and contingent response repetition) are used to facilitate the initial appearances of accurate responses. When responding is consistently accurate, practice (i.e., frequent and repeated opportunities to respond) promotes response fluency. Performance feedback for rate of responding is also likely to improve fluency.

Several studies have demonstrated the utility of these distinctions by finding individual differences in students' responsiveness to performance-based (i.e., programmed reinforcement) or skill-based (i.e., use of instruction) interventions (Duhon et al., 2004; Eckert, Ardoin, Daisey, & Scarola, 2000; Eckert, Ardoin, Daly, & Martens, 2002; Noell et al., 1998). For example, Duhon et al. (2004) used brief assessment procedures to generate hypotheses about skill versus performance deficits in the areas of math and writing. Skill deficits were displayed by two students and performance deficits were displayed by two students. The hypotheses were confirmed through extended classroom applications of both types of interventions for all four students. In all cases, the results validated the original hypotheses which were formulated based on student assessments.

Establishing functional relationships for generalized word reading.

Experimental analyses that have explicitly targeted generalization of word reading by manipulating word overlap have been few in number. Many studies have measured outcomes directly in training materials, testing for mastery but not for generalization. One exception is a study by Daly, Martens, et al. (1996), who found an interaction between word overlap and difficulty level. The greatest effects were achieved when there was high word overlap between instructional and assessment conditions and difficulty level was better matched to students' instructional level (i.e., the materials were not too hard). Since that time, several studies have incorporated high word overlap passages into the experimental analyses, many of which will be reviewed in a later section. Here, we wish to focus on the results of a study that examined the effects of a combination of instructional and motivational variables on generalization to high word overlap passages. Daly, Bonfiglio, Mattson, Persampieri, and Yates (in press) found that a combination of antecedent instructional variables (in an instructional passage) and reinforcement for generalized responding (in high word overlap passages) produced greater generalization than a reinforcement-only condition, suggesting that it may be necessary to combine instructional and reinforcement components to produce generalized word reading in some cases. Generalization of responding was directly reinforced (across stories with different word order) and the instructional antecedents to generalization appear to have increased the probability of correct responding during the assessment of generalization. It is possible that the contingency which was explained at the beginning of the session may have increased motivation to attend to the discriminations being taught (i.e., word reading fluency) during instruction. Interestingly, there was an interaction effect with difficulty level, with two participants displaying higher relative increases in easier passages and one participant displaying higher relative increases in harder passages. Individual differences in student responding to instructional and/or motivational components signals a need to test intervention components prior to instructional intervention if maximum impact is sought.

The interventions that lead to stimulus control and stimulus generalization are obviously important. Individual differences between students and the necessity of efficient interventions that are not costly in terms of time and effort in schools speak to the need to identify which intervention components may be necessary for a particular student. A number of strategies are available to the practitioner. Strategies used in the experimental analyses described in this paper appear in Table 1. These are the intervention components that we have found to be particularly useful for establishing stimulus control and generalization. They have been used together as a single treatment package and in various combinations to establish stimulus control and stimulus generalization for word reading (as will be reviewed in the next section).

Table 1. *Reading Fluency Intervention Components*

Component	Rationale	Procedural Steps
Reward (R)	Used to identify performance deficits (Daly, Murdoch, Lillenstein, Webber, & Lentz, 2002)	The practitioner tells the student that a tangible item (e.g., bouncy balls, pencils, stickers, candy, etc.) or access to a privilege (e.g., 10 min of playing a game) is available to the student contingent upon meeting a predetermined individualized performance goal. The performance goal is based on a 30% increase in correct words per minute, with fewer than 4 errors, derived from the student's previous performance on the passage. Prior to instruction and assessment, the student chooses one reward to earn for meeting the goal. The reward is delivered after the assessment if the student met or exceeded the goal. This condition can be used to reward generalization of responding if prior instruction is carried out in an instructional passage that has high word overlap.
Listening Passage Preview (LPP)	Provides modeling to increase the student's reading accuracy and fluency (Daly & Martens, 1994).	The examiner reads the instructional passage to the student at a comfortable pace while simultaneously monitoring the student to ensure that he or

		she is correctly following along with his or her finger.
Repeated Readings with Performance Feedback (RR)	Designed to provide a student with multiple opportunities to respond by having the student re-read a passage repeatedly and provide feedback on fluency (Eckert et al., 2002)	The examiner has the student re-read a passage three times. Each time, the examiner tells the student how quickly he or she read the passage and how many errors were made.
Phrase Drill (PD)	Designed to provide corrective feedback and accurate practice to increase correct responding (O'Shea, Munson, & O'Shea, 1984)	As the student reads the passage the first time, the practitioner highlights or underlines the student's errors. After the student finishes reading the passage, the practitioner points to and reads the first error word to the student. The student reads the error word correctly to the practitioner, and then reads the phrase or sentence containing the error word three times. This process is repeated for each error word.
Syllable Segmentation (SS)	Designed to increase accuracy by providing the students with further corrective feedback and practice blending the syllables of error words together (Daly et al., in press).	After the student has read the passage a second time, the practitioner corrects errors by using an index card to cover each error word and uncovering and modeling the correct pronunciation of one syllable at a time. The student repeats the correct pronunciation of each syllable as the practitioner uncovers them. The student then independently reads each syllable and blends the syllables together to pronounce the word. If the student makes any mistakes during this process, the practitioner repeats the previous step.

Identifying Oral Reading Fluency Interventions Through Brief Experimental Analysis

Experimental analysis has a long, honorable, and fruitful tradition as the analytic framework within which empirical validation has occurred in the field of applied behavior analysis (Johnston & Pennypacker, 1990; Sidman, 1960). Unfortunately, the field has been slow to expand experimental analysis beyond social behaviors (Ervin et al., 2001). Recently, however, experimental analysis has begun to be used as a methodology for identifying effective treatment conditions, much as functional analysis was developed to facilitate treatment selection for behavioral excesses (Daly, Witt, Martens, & Dool, 1997). At least three characteristics of brief experimental analysis for academic behaviors differentiate it from traditional functional analyses. First, the analyses are conducted with behavioral deficits (i.e., not enough academic responding) rather than with behavioral excesses (e.g., self-injurious behavior). Second, treatments are applied directly as a part of the experimental analysis (and not inferred based on analyses of maintaining variables). Third, data series and conditions are often abridged to make the process more time efficient (Martens, Eckert, Bradley, & Ardoin, 1999). It is for this last reason that the methods and procedures have inherited the name "brief experimental analysis," or BEA, for short.

Applications of BEA. BEA has been used to generate effective reading interventions for parent tutoring (Daly, Shroder, & Robinson, 2001; Gortmaker, Daly, McCurdy, Persampieri, & Hergenrader, 2005; Persampieri, Gortmaker, Daly, & Sheridan, in press; Valleley, Evans, & Allen, 2002), small reading groups (Bonfiglio, Daly, Persampieri, & Andersen, 2005), and self-managed interventions (Daly, Persampieri, McCurdy, & Gortmaker, 2005). When used to develop parent tutoring interventions, it has potential for maximizing treatment integrity when it identifies the intervention that yields the best results yet requires the least amount of effort (Valleley et al., 2002). When the parent conducts an instructional trial as a part of the BEA, he or she not only gets to "try out" the intervention, but also receives supervision and training from the one supervising the BEA (Persampieri et al., in press). Results can be compared to those obtained by the clinician. The same is true for applications to small reading groups when the teacher uses the indicated treatment as a last step before classroom application (Bonfiglio et al., 2005). Finally, many of the instructional components used as a part of BEA can be tailored to individualized, self-managed components that require a minimum of adult supervision when the classroom teacher is unable or unwilling to modify typical reading instruction (Daly et al., in 2005). There are other potential applications of BEA derived interventions that have not yet been explored in the literature (e.g., peer tutoring). However, these examples illustrate that application of BEA results can be accomplished in a variety of ways.

Three methods for conducting BEAs. Since its development, three approaches to designing BEAs for reading fluency problems have been taken. Early on, intervention components were evaluated singly (Daly, Martens, Dool, & Hintze, 1998; Jones & Wickstrom, 2002; Vallely et al., 2002). For example, Daly et al. applied intervention components individually until a visible increase in reading fluency was found. Once this increase was obtained, the investigators added an additional baseline and then re-introduced the effective intervention component for experimental control purposes. Replication of baseline and the effective condition strengthened the case for the selected intervention. By only introducing an additional baseline condition when the intervention was identified as effective, Daly et al. reduced the overall number of sessions needed. However, the evaluation can still take a number of sessions to identify a single instructional component that produces a strong effect.

Although combining treatment components may create more complex treatments, effects would probably be stronger and may more closely resemble actual classroom instruction. Teachers would rarely (if ever) use a single instructional technique only. Therefore, BEAs in which intervention components were added sequentially began to emerge (Daly, Martens, Hamler, Dool, & Eckert, 1999; Daly, Murdoch, Lillenstein, Webber, & Lentz, 2002; VanAuken, Chafouleas, Bradley, & Martens, 2002). For example, Daly et al. (2002) systematically evaluated combinations of repeated readings, listening passage preview, phrase drill error correction, sequential modification (a generalization strategy), text difficulty, word list training, and rewards to identify effective interventions for five second grade students. Following a baseline condition, intervention began with a single component (repeated readings) and proceeded sequentially by including an additional treatment component in each subsequent condition. Individual differences were obtained in students' responsiveness to the treatment combinations, with some students requiring simpler and some more complex treatments.

The third approach that has been taken to conduct BEAs is to use a strong treatment package initially and dismantle the package until the simplest intervention that still produces reasonable increases in performance is identified. For example, Daly et al. (2005) conducted brief experimental analyses in three phases. The first phase included a treatment package consisting of both skill- and performance-based strategies at two difficulty levels (i.e., easier and harder) and control conditions. Then, the package was dismantled by separating skill-based and performance-based instructional components. Finally, the indicated treatment (reinforcement-only for one student and the treatment package for the other) was compared once again to control and another treatment for validation purposes. Intervention components were implemented by the students with the assistance of the experimenter in extended analyses and led to substantial increases in reading performance for both students. The advantage to the dismantling approach is that the complete treatment package can serve as a benchmark against which leaner treatments can be compared. If a simpler intervention produces the same result as the treatment package, then the simpler intervention is recommended for adoption as the intervention of choice.

Conducting a Single Instructional Trial BEA

In this section, guidelines for conducting a BEA are presented. These procedures can be used to identify an intervention in a single instructional trial. They are based on methods used in the studies described earlier, but have been simplified to reduce the amount of time and number of sessions necessary to identify an appropriate intervention. Student performance is measured immediately after the instructional trial in three different passages, allowing the examiner to determine whether generalization gains have been made as a function of either a treatment package (containing both instructional and reward components) and/or a reward-only condition relative to a control condition. An initial screening is conducted that should take not more than about 15 minutes. The instructional and assessment session can be conducted in about 20 minutes. The steps are presented in Table 2. After explaining how to prepare for a BEA, an explanation of each step is given.

Table 2

Steps for Conducting a Single Instructional Trial BEA

Steps
1. Screen to identify at least three equal difficulty level assessment passages.
2. Randomly assign one passage to the treatment package, one to the reward-only condition, and one to the control condition.
3. Deliver the treatment package using the corresponding instructional passage for the assessment passage assigned to the treatment package condition.
4. Assess student performance in all three passages immediately after treatment. Order of passages should be randomized.
5. Reward is delivered contingent on meeting pre-specified criteria for performance in one of the two reward passages (reward-only and treatment package). The passage chosen for reward is determined randomly between the two options after the student has read all three passages.

Materials preparation. Two types of reading passages are used in a BEA: assessment and instructional passages. All passages should consist of short (i.e., approximately 150 words) stories at the level at which the student is currently being instructed. Each assessment passage should have a corresponding high word overlap instructional passage. Assessment passages are rewritten as a different story with a high percentage of the same words to create instructional passages. Passages used in our research and practice have generally had about 80 to 95% word overlap. The percentage of word overlap can be calculated by dividing the number of words contained in both passages by the total number of words in the assessment passage. Each assessment passage is used twice in the BEA—first during screening to identify equal difficulty level passages, and then again after the student has received the instructional trial in an instructional passage.

Although each assessment passage has a corresponding high word overlap instructional passage, only one of the instructional passages will be used during the analysis (the one randomly chosen for the treatment package following screening). Therefore, one assessment passage will have high word overlap with the instructional passage: this is the treatment package passage. Two assessment passages will have low word overlap with the instructional passage: these are the control and reward passages. We recommend having at least a dozen assessment passages on hand for the screening to increase the likelihood of finding three equal difficulty level passages during the screening. You will also need two flashcards, one with an “A” marked on one side and the other with a “B” marked on one side. (Be sure that the marking is not visible from the other side.) These flashcards will be used to determine to which passage the reward criterion will be applied.

Pre-experimental screening. The purpose of the pre-experimental screening is to identify equal difficulty level passages that will serve as the basis for comparing conditions. Equating difficulty level controls for variance in reading performance due to fluctuations in passage difficulty from one passage to another. (Our experience has led us to find that readability formulas do a poor job of reflecting difficulty level of a passage for a given student and that the best way to determine difficulty level is to measure the student's oral reading fluency performance in the passage.) Once the materials are gathered, administer all of the high word overlap *assessment* passages to the student for 1 minute each in random order to collect the baseline oral reading fluency for each passage (CRW per min and errors per min according to standard CBM administration procedures; Shinn, 1989).

Pre-session preparation. Once all of the baseline fluency scores for the passages are collected, sort them from highest to lowest based on CRW per min. For example, if a student reads 46, 48, 33, 31, 37, 34 and 35 CRW per min on a series of passages, sort the scores as 48, 46, 37, 35, 34, 33 and 31 CRW per min. Next, choose the three passages that are closest in difficulty level (i.e., the passages for which the student read 35, 34, and 33 CRW per min in the example). Randomly assign one passage to the treatment package condition, one to the control condition, and one to the reward condition. Student copies of the two passages in which the student can earn a reward should be indicated in some way (e.g., with "REWARD" written across the top). The criterion for meeting the reward should be indicated on the examiner copies for these two passages. We recommend a criterion of a 30% increase in performance over the screening results for the passage with 3 or fewer errors (Daly et al., 2005). For example, for the passage in which the student read 34 CRW per min during screening, a 30% improvement would be 44 CRW per min. The criterion is determined individually for each passage. We also suggest that you put a bracket after the last word in the examiner copies of the two passages that must be met for the student to earn a reward (but not in the student copy). Differentiate the passages from one another by marking one passage as "REWARD A" and the other as "REWARD B" or some such other designation.

Conducting the instructional trial.

The high word overlap instructional passage that is associated with the assessment passage assigned to the full treatment package is selected for the instructional trial. All of the treatment components except the reward condition (see Table 1) are administered to the student in this passage. However, before beginning the instructional trial, the examiner allows the student to choose a reward (e.g., a tangible, access to a privilege, an edible) toward which he or she will be able to work during the assessment passages. Student motivation may be increased if the examiner explains that the instructional passage has a lot of the words as one of the assessment passages and that practicing in the instructional passage may help him or her to do well in the assessment passage. The steps for conducting the instructional trial are outlined in Table 3.

Table 3. Protocol for the Instructional Trial

Steps
1. Explain the reward contingency that will be applied to the two reward/assessment passages. Explain to the student that practicing in the instructional passage may help him or her do better in the reward passages.
2. Taking the instructional passage, read the passage aloud to the student at a comfortable reading rate while he or she follows along with a finger.
3. Have the student read the passage for 2 minutes while you mark errors. When the student is done, tell him or her how fast he or she read the passage (CRW per min) and how many errors he or she made.
4. Read each error word to the student and have him or her read the sentence containing the error word three times. Model correct responding if the student continues to make errors.
5. Have the student read the passage a second time for 2 minutes while you mark errors. When the

student is done, tell him or her how fast he or she read the passage (CRW per min) and how many errors he or she made.

6. For words that were read incorrectly a second time (i.e., were read incorrectly during both steps 3 & 5), break each word into syllables for the student. Next, have the student break the words into syllables (repeating what you just did) and also blend the syllables together to form the correct word. Model correct responding if the student continues to make errors.

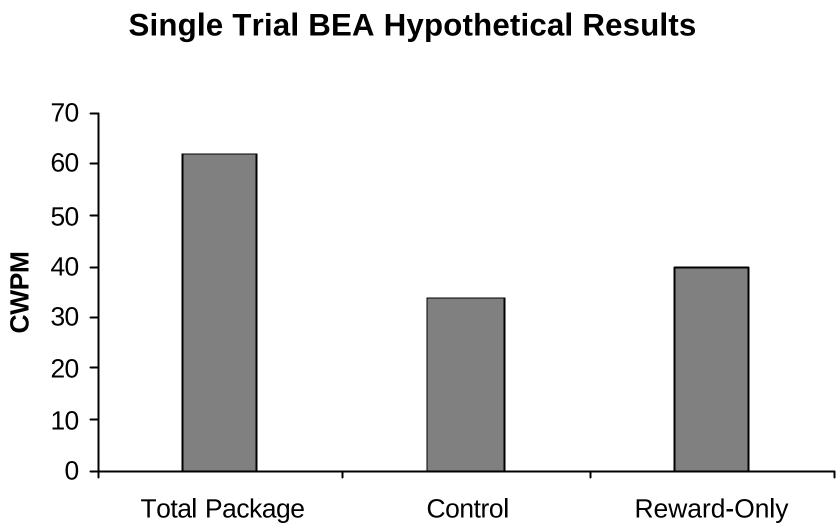
Assessing performance and giving feedback about reward. Before assessment begins, the examiner explains to the student that he or she can earn a reward for beating his or her last score on one of the two passages marked with "REWARD" at the top, and that the examiner will tell the student whether he or she met the goal after all three passages have been administered. The three assessment passages are administered to the student for 1 minute each in random order while the examiner scores student performance for CRW per min and errors in each passage. We suggest that you signal to the student each time a reward passage is presented (e.g., "You can see that this is one of the reward passages.").

At the end of the session, set the passages aside and explain to the student that you will determine together which passage is the reward passage. Use the following procedure to randomly choose the passage to which the contingency will be applied. Shuffle the two index cards and present them with the blank sides facing the student so that he or she cannot see which card is which. Have the student choose a card without knowing whether it is "A" or "B." Then, determine whether the student met the goal in the passage indicated by the flashcard (e.g., passage "A"). If the student met the criterion for performance (in terms of cumber of CRW per min and errors), offer access to the reward.

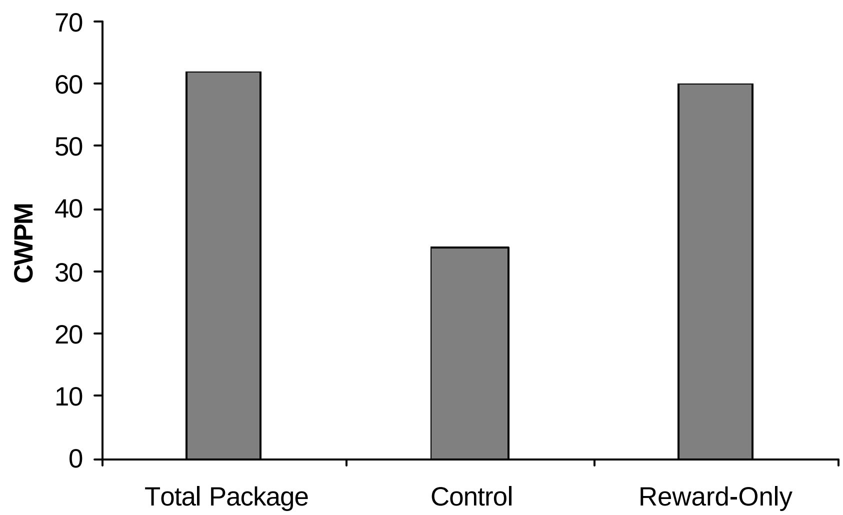
Interpretation of results. Results should be plotted on a graph. Figure 1 depicts three possible outcomes of the BEA. If the treatment package results exceed the other conditions (top graph in the Figure), then the evaluator can decide either to move on to treatment implementation or attempt to dismantle the treatment further (see next section). If reward meets or exceeds results of the treatment package and both exceed the control passage (middle graph in the Figure), then the student has a performance deficit which can be managed through rewards for performance gains. If the student fails to increase in either of the two conditions relative to the control condition (bottom graph in the Figure), the practitioner should consider either moving down to easier materials (VanAuken et al., 2002) or applying instructional components to the assessment passages as well (sequential modification; Daly et al., 1999). In either case, the student has a significant generalization problem and will need a very intensive intervention. (See Daly et al., 2002, for other instructional components that can be tried.)

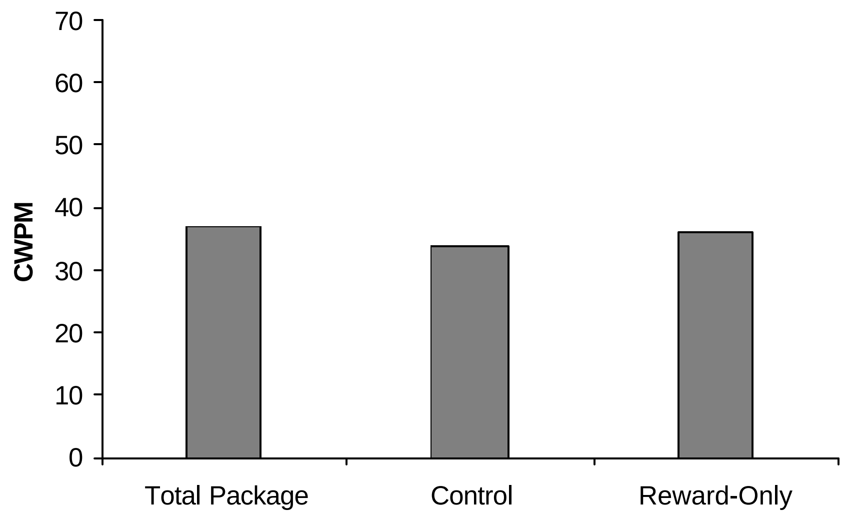
FIGURE 1, NEXT PAGE!

Figure 1. Hypothetical results for a single trial brief experimental



analysis.





Optional steps for further dismantling the instructional condition.

Upon visual inspection of the results of the first instructional trial, if the student performs notably higher in the treatment package passage than the reward-only and control passages, the student probably has a skill deficit. The evaluator can do one of two things at this point. Either, the assessment can be terminated because a treatment has been identified (i.e., the treatment package that includes both instructional and motivational components). Alternately, the instructional package can be dismantled further by sequentially withdrawing instructional components until the most effective, yet simplest instructional package is identified. This step can be taken if the evaluator is concerned that the person responsible for implementing the intervention (e.g., a parent, teacher, peer tutor) may not be able to follow all of the steps of the instructional protocol consistently.

If the decision is made to dismantle the instruction package further, further screening will need to be conducted to identify more passages for the analysis. As many as six additional, equal difficulty level assessment passages may be needed. The examiner should then proceed with the following conditions until there is a clear drop in performance. Each of these instructional conditions is carried out in the instructional passage for one assessment passage. Assessment should be carried out in random order in the assessment passage and an equal difficulty level control passage for each session. First, the examiner should administer the full instructional treatment without the reward. If performance matches or exceeds previous performance, the examiner should then withdraw the error correction components (phrase drill & syllable segmenting, leaving listening passage preview and repeated readings) in the next session (and set of passages) because they provide the fewest opportunities to respond (Gortmaker et al., 2005). If performance does not drop, withdraw the LPP component and administer the RR component in the next session (and assigned instructional passage). If performance does not drop, one concludes that the student will probably benefit from the RR intervention. A drop in performance in any of these conditions indicates that the previous instructional trial contained critical instructional or motivational components that are necessary for improving student performance.

Limitations to BEA

With the rapid alternation of multiple treatments implemented in brief conditions, the risk of multiple treatment interference naturally arises as a threat to the internal validity of an analysis. This threat is minimized with the use of a control passage in each condition and replication of the chosen intervention during the BEA. Furthermore, investigations integrating extended analyses provide support for the effectiveness of interventions derived from BEA (Daly et al., 2005; Daly et al., 2001; Gortmaker et al., 2005; Jones & Wickstrom, 2002; Persampieri et al., in press; Valleley et al., 2002).

Research to date on the BEA of academic performance has largely focused on reading fluency. The procedures described in this article are applicable mostly to students who are able to read with some degree of fluency in text and are less appropriate for non-readers. Although this form of analysis has been applied to reading comprehension, spelling, math, and writing (Daly et al., 1998; Duhon et al., 2004; Hendrickson, Gable, Novak, & Peck, 1996; Jones & Wickstrom, 2002; McComas et al., 1996; Noell et al., 1998; VanAuken et al., 2002), more research is needed on its application to these and other early literacy skills like phoneme blending and segmenting.

Conclusion

It was noted earlier in this paper that general outcome measurement (formative evaluation) is the strongest measurement model available to educators. Although BEA may help educators to identify an intervention in an efficient manner, it does not eliminate the need to monitor student progress over time and make instructional adjustments accordingly. Educators should use materials like those available as a part of Dynamic Indicators of Basic Early Literacy Skills assessments (DIBELS; Good & Kaminski, 2002) or through Aimsweb® reading series (Edformation, 2005) for the purpose of monitoring students' generalization to materials that have not been directly taught in the classroom. These are the types of generalized improvements in basic skills like oral reading fluency that will make it easier for students to move on to harder parts of the curriculum (Binder, 1996). We propose BEA merely as an intermediary form of connecting the dots.

References

Alessi, G. (1987). Generative strategies and teaching for generalization. *The Analysis of Verbal Behavior, 5,* 15-27.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19,* 163-197.

Bonfiglio, C. M., Daly, E. J., III, Persampieri, M., Andersen, M. N. (2005). An experimental analysis of the effects of reading interventions in a small group reading instruction context. Submitted for publication, *Journal of Behavioral Education.*

Catania, A. C. (1998). *Learning* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Cunningham, A., Stanovich, K. (1998). The impact of print exposure on word recognition. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 235-262). Mahwah, NJ: Lawrence Erlbaum Associates.

Daly, E. J., III, Bonfiglio, C. M., Hauger, T., Persampieri, M., & Yates, K. (in press). Refining the experimental analysis of academic skill deficits, Part II: An Investigation of the use of brief experimental analysis for identifying reading fluency interventions. *Journal of Applied Behavior Analysis.*

Daly, E. J., III, Lentz, F. E., & 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., III, & 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.

Daly, E. J., III, 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., III, 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., III, Martens, B. K., Kilmer, A., & Massie, D. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29,* 507-518.

Daly, E. J., III, Murdoch, A., Lillenstein, L., Webber, L., & Lentz, F. E. (2002). An examination of methods for testing treatments: Conducting experimental analyses of the effects of instructional components on oral reading fluency. *Education and Treatment of Children, 25,* 288-316.

Daly, E. J., III, Persampieri, M., McCurdy, M., & Gortmaker, V. (2005) Generating reading interventions through experimental analysis of academic skills: Demonstration and empirical evaluation. *School Psychology Review, 34,* 395-414.

Daly, E. J., III, Shroder, J., & Robinson, A. (2001). Testing treatments for oral reading fluency problems: Two case studies. *Proven Practice: Prevention and Remediation Solutions for Schools, 4,* 2-10.

Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26,* 554-574.

Deno, S. L., Fuchs, L. S., Marston, D., & Jongho, S. (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Duhon, G. J., Noell, G. H., Witt, J. C. Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skills and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429-443.

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: An examination of the efficacy of combining skill-based and performance-based interventions. *Journal of Applied Behavior Analysis, 35,* 271-281.

Edformation. (2005). *Aimsweb: Charting the path to literacy.* (Available on-line at http://www.aimsweb.com/)

Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, A. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review, 30,* 193-210

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28,* 659-671.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488-500.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Effects of instructional use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43-52.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review, 20,* 49-66.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

Fuchs, L., Shinn, M. (1989). Writing CBM IEP objectives. In M Shinn (Ed.) *Curriculum-based measurement: Assessing special children* (pp. 130-152). New York: Guilford Press.

Good, R. H., III, & Kaminski, R. A. (2002). *DIBELS oral reading fluency passages for first through third grades.* (Technical Report No. 10), Eugene, OR: University of Oregon.

Gortmaker, V. J., Daly, E. J., III, McCurdy, M., Persampieri, M. J., & Hergenrader, M. (2005). *Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent tutoring interventions.* Manuscript submitted for publication.

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

Hendrickson, J. M., Gable, R. A., Novak, C., & Peck, S. (1996). Functional assessment as strategy assessment for teaching academics. *Education and Treatment of Children, 19,* 257-271.

Hintze, J. M., Daly, E. J., III, & Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27,* 433-445.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.

Jones, K. M., & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31,* 554-568.

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

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.

Meyer, M., Felton, R. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49,* 283-306.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* Available on-line: http://www.nichd.nih.gov/publications/nrp/smallbook.htm

Noell, G. H., Gansle, K. A., Witt, J. D., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., Gilbertson, D. N., & Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659-664.

O'Shea, L. J., Munson, S. M., & O'Shea, D. J. (1984). Error correction in oral reading: Evaluating the effectiveness of three procedures. *Education and Treatment of Children, 7*, 203-214.

Persampieri, M., Gortmaker, V., Daly, E.J., III, Sheridan, S. M., McCurdy, M. (in press). Promoting parent use of empirically supported reading interventions: Two experimental investigations of child outcomes. *Behavioral Interventions.*

Shahan, T. A., & Chase, P. N. (2002). Novelty, stimulus control, and operant variability. *The Behavior Analyst, 25,* 175-190.

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

Shinn, M. R. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 547-567). Washington, DC: National Association of School Psychologists.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: The National Research Council.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-367.

Valleley, R. J., Evans, J. H., & Allen, K. D. (2002). Parent implementation of an oral reading intervention: A case study. *Child & Family Behavior Therapy*, *24*, 39-50.

VanAuken, T., Chafouleas, S.M., Bradley, T.A., & Martens, B.K. (2002). Using brief experimental analysis to select oral reading interventions: An investigation of treatment utility. *Journal of Behavioral Education*, *11*, 165-181.

Authors' Contact Information:

Edward J. Daly III
University of Nebraska-Lincoln
33 Teachers College
Lincoln, NE 68588-0345
(402) 472-5923
edaly2@unl.edu.

Melissa Andersen
University of Nebraska-Lincoln
35 Teachers College
Lincoln, NE 68588-0345
(402) 472-2207
melissandersen@gmail.com.

Valerie Gortmaker
University of Nebraska-Lincoln &
University of Nebraska Medical Center
Psychology Department
Munroe-Meyer Institute
985450 Nebraska Medical Center
Omaha, NE 68198-5450
valeriegortmaker@yahoo.com.

April Turner
University of Nebraska-Lincoln
35 Teachers College
Lincoln, NE 68588-0345
(402) 472-2207
april_d_turner@yahoo.com.