Education and Training in Developmental Disabilities

Focusing on individuals with
cognitive disabilities/mental retardation, autism, and related disabilities

Volume 40 Number 4 DDD December 2005

December 2005 Education and Training in Developmental Disabilities Vol. 40, No. 4, pp. 333–428

Education and Training in Developmental Disabilities

The Journal of the Division on Developmental Disabilities, The Council for Exceptional Children

Education and Training in Developmental Disabilities is sent to all members of the Division on Developmental Disabilities of The Council for Exceptional Children. All Division members must first be members of The Council for Exceptional Children. Division membership dues are $25.00 for regular members and $8.00 for full time students. Membership is on a yearly basis. All inquiries concerning membership, subscription, advertising, etc. should be sent to the Division on Developmental Disabilities, 1110 North Glebe Road, Arlington, VA 22201. Advertising rates are available upon request.

Manuscripts should be typed, double spaced, and sent (five copies) to the Editor: Stanley H. Zucker, Special Education Program, Box 872011, Arizona State University, Tempe, AZ 85287-2011. Each manuscript should have a cover sheet that gives the names, affiliations, and complete addresses of all authors.

Editing policies are based on the **Publication Manual,** the American Psychological Association, 2001 revision. **Additional information is provided on the inside back cover.** Any signed article is the personal expression of the author; likewise, any advertisement is the responsibility of the advertiser. Neither necessarily carries Division endorsement unless specifically set forth by adopted resolution.

Education and Training in Developmental Disabilities is abstracted and indexed in **Psychological Abstracts, PsycINFO, e-psyche, Abstracts for Social Workers, International Journal of Rehabilitation Research, Current Contents/Social and Behavioral Sciences, Excerpta Medica, Social Sciences Citation Index, Adolescent Mental Health Abstracts, Educational Administration Abstracts, Educational Research Abstracts,** and **Language and Language Behavior Abstracts.** Additionally, it is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly print index **Current Index to Journals in Education** and the quarterly index, **Exceptional Child Education Resources.**

Division on Developmental Disabilities

The Council for Exceptional Children

The purposes of this organization shall be to advance the education and welfare of persons with developmental disabilities, research in the education of persons with developmental disabilities, competency of educators in this field, public understanding of developmental disabilities, and legislation needed to help accomplish these goals. The Division shall encourage and promote professional growth, research, and the dissemination and utilization of research findings.

EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES (ISSN 1547-0350) (USPS 0168-5000) is published quarterly in March, June, September, and December, by The Council for Exceptional Children, Division on Developmental Disabilities, 1110 North Glebe Road, Arlington, Virginia 22201-5704. Members' dues to The Council for Exceptional Children Division on Developmental Disabilities include $8.00 for subscription to EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES. Subscription to EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES is available without membership; Individual—U.S. $30.00 per year; Canada, PUAS, and all other countries $34.00; Institutions—U.S. $75.00 per year; Canada, PUAS, and all other countries $79.50; single copy price is $20.00. U.S. Periodicals postage is paid at Arlington, Virginia 22204 and additional mailing offices.

POSTMASTERS: Send address changes to EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES, 1110 North Glebe Road, Arlington, Virginia 22201-5704.

Education and Training in Developmental Disabilities

Editorial Policy

Education and Training in Developmental Disabilities focuses on the education and welfare of persons with developmental disabilities. *ETDD* invites research and expository manuscripts and critical review of the literature. Major emphasis is on identification and assessment, educational programming, characteristics, training of instructional personnel, habilitation, prevention, community understanding and provisions, and legislation.

Each manuscript is evaluated anonymously by three reviewers. Criteria for acceptance include the following: relevance, reader interest, quality, applicability, contribution to the field, and economy and smoothness of expression. The review process requires two to four months.

Viewpoints expressed are those of the authors and do not necessarily conform to positions of the editors or of the officers of the Division.

Submission of Manuscripts

1. Manuscript submission is a representation that the manuscript is the author's own work, has not been published, and is not currently under consideration for publication elsewhere.
2. Manuscripts must be prepared according to the recommendations in the **Publication Manual of the American Psychological Association** (Fifth Edition, 2001). Standard typewriter type, laser, or high density dot printing are acceptable.
3. Each manuscript must have a cover sheet giving the names and affiliations of all authors and the address of the principal author.
4. Graphs and figures should be originals or sharp, high quality photographic prints suitable, if necessary, for a 50% reduction in size.
5. Five copies of the manuscript along with a transmittal letter should be sent to the Editor: Stanley H. Zucker, Special Education Program, Box 872011, Arizona State University, Tempe, AZ 85287-2011.
6. Upon receipt, each manuscript will be screened by the editor. Appropriate manuscripts will then be sent to consulting editors. Principal authors will receive notification of receipt of manuscript.
7. The Editor reserves the right to make minor editorial changes which do not materially affect the meaning of the text.
8. Manuscripts are the property of **ETDD** for a minimum period of six months. All articles accepted for publication are copyrighted in the name of the Division on Developmental Disabilities.

Education and Training in Developmental Disabilities

VOLUME 40 NUMBER 4 DECEMBER 2005

Manuscripts Accepted for Future Publication in Education and Training in Developmental Disabilities

March 2006

Attitudes of Japanese adults towards persons with intellectual disability: Effect of perceptions concerning intellectual disability **Toshiaki Tachibana,** Department of Social Welfare, Aichi Shinshiro Otani University, Kawaji, Shinshiro, Aichi 441-1306 JAPAN.

Correspondence between video CD-ROM and community-based job preferences for individuals with developmental disabilities. David A. Ellerd, **Robert Morgan,** and Charles L. Salzberg, Utah State University, 2865 Old Main, Logan, UT 84322-2865.

Keeping them happy: Job satisfaction, personality, and attitudes toward disability in predicting counselor job retention. Emily R. Lawrence, **Laraine Glidden,** and Brian M. Jobe, St. Mary's College of Maryland, 18952 East Fisher Road, St. Mary's City, MD 20686-3001.

Active and passive task related behavior, direction following and the inclusion of children with disabilities. **Coral Kemp** and Mark Carter, Macquarie University Special Education Centre, Macquarie University, Sydney, 2109 AUSTRALIA.

Infusing self-determination into 18-21 services for students with intellectual or developmental disabilities: A multi-stage, multiple component model. **Michael L. Wehmeyer,** Nancy Garner, Danna Yeager, Margaret Lawrence, and Anna Kay Davis, University of Kansas, Haworth Hall, 1200 Sunnyside Ave., Room 3136, Lawrence, KS 66045-7534.

Function-based intervention to support the inclusive placements of young children in Korea. Kwang-Sun Cho Blair, **Carl J. Liaupsin,** John Umbreit, and Gilbok Kweon, University of Arizona, College of Education Room 412, PO Box 210069, Tucson, AZ 85721.

Teaching special education teachers how to conduct functional analysis in natural settings. **Dilek Erbas,** Elif Tekin-Iftar, and Serife Yucesoy, Anadolu Universitesi, DILKOM, Eskisehir 26470 TURKEY.

Predicting poor achievement in early grade school using kindergarten scores on simple cognitive tasks. **Marcia Strong Scott** and Christine F. Delgado, University of Miami, Department of Psychology, Flipse Building, P.O. Box 248185, Coral Gables, FL 33124-0751.

Address is supplied for author in boldface type.

Education and Training in Developmental Disabilities, 2005, 40(4), 335–342

Self-Injurious Behavior and Functional Analysis: Ethics and Evidence

Richard P. Hastings and Stephen J. Noone
University of Wales Bangor, UK

Abstract: Ethical guidelines for behavior analysts state that functional assessment should be used before implementing treatment plans. In this paper, we discuss the ethical implications of this position using self-injurious behavior in people with developmental disabilities as a focus. Evidence is reviewed that suggests that treatment based on results of a functional assessment is likely to be more effective, less likely to be considered restrictive, more likely to be perceived as socially valid, and also focuses on principles of a constructional approach to behavior change. Alongside this generally positive evidence, there are ethical problems with the validity of some assessment methods, their potential lack of clarity, and the intensive resources required by a functional assessment. Effective, socially valid, and constructional interventions for self-injury might also be developed using preventative interventions. The balance of all of these ethical issues needs to be considered in the design of any treatments for self-injurious behavior.

The Behavior Analyst Certification Board (BACB) *Guidelines for Responsible Conduct for Behavior Analysts* (BACB, 2001) state clearly that behavior analysts should conduct functional assessment as a part of their work with a client. The purpose of the present paper is to explore the ethical implications of conducting or not conducting a functional assessment. The discussion will focus on the case of self-injurious behavior in people with developmental disabilities. Self-injury is a major clinical problem that can result in significant injury and even death, and has been associated with negative outcomes for carers such as stress (Hastings, 2002). Epidemiological surveys suggest that up to 20% of adults with mental retardation engage in some form of self-injurious behavior (Collacott, Cooper, Branford, & McGrother, 1998; Emerson et al., 2001; Oliver, Murphy, & Corbett, 1987). Thus, self-injury is a problem of considerable applied significance falling within the purview of Applied Behavior Analysis (ABA; Baer, Wolf, & Risley, 1968).

Correspondence concerning this article should be addressed to Richard Hastings, PhD, School of Psychology, University of Wales Bangor, Bangor, Gwynedd, LL57 2AS, UK. E-mail: r.hastings@bangor.ac.uk

There are a number of principles applicable to the question of whether treating self-injury with an intervention based on a pre-treatment functional assessment is ethically superior to implementing an intervention without a functional assessment. Those that will be discussed here are: the right to effective treatment, the application of least restrictive treatment alternatives, social validity, and the importance of a constructional approach within ABA. Before discussing these ethical issues, it is important to define some terminology to be used in this paper. 'Functional Assessment' will be used to refer to a whole variety of methods that might be used to inform hypotheses about the function of self-injury, whereas 'Analog Assessment' will be reserved specifically to refer to experimental functional analysis (cf. Iwata, Dorsey, Slifer, Bauman, & Richman, 1982). Some of the ethical issues relating to experimental and non-experimental methods of functional assessment are different and so this is an important distinction to maintain.

The contrast in the discussion below is between treatment based on functional assessment, and an approach referred to as 'Behavior Modification'. Behavior Modification is defined here as the application of behavior change technologies (i.e., via reinforcement

or punishment techniques) without reference to the analytic dimension of ABA (Baer et al., 1968). One of the defining features of ABA is that it is analytic in nature: involving a reliable demonstration of the events responsible for the occurrence and non-occurrence of behavior (Baer et al.). This central distinction of whether treatments are based on hypotheses about their causes (Repp, Felce, & Barton, 1988) has ethical implications for the treatment of self-injury in individuals with developmental disabilities.

Effective Treatment for Self-Injury

Effectiveness of treatment can be defined in a number of ways. At the basic level, it is certainly possible to identify reports in the research literature demonstrating that behavior modification procedures for self-injury can lead to clinically salient outcomes (see Repp, Singh, Olinger, & Olson, 1990, for a review). Similarly, there are reports demonstrating that clinically salient outcomes are achieved by carrying out a functional assessment and deriving a treatment for self-injury from hypotheses about the processes maintaining the behaviour (e.g., Carr & Durand, 1985). The central ethical point is that clients should be offered the most effective treatments available.

Probably the most convincing evidence that addresses the question of relative effectiveness comes from research employing meta-analytic methods to explore variables that moderate the effectiveness of treatment for severe behavior problems such as self-injury. In reviewing a large number of single case experiments reporting treatment of severe behavior problems, several authors have used a standard statistic called the Percentage of Non-Overlapping Data points (PND) (Scruggs & Mastropieri, 1998) to quantify treatment effects within each experiment. A very effective treatment will have a PND of 1.0 (i.e., there is no overlap between data points during baseline and treatment phases). Further characteristics of each experiment can be coded (e.g., age of client, severity of disability, whether a pretreatment functional assessment was conducted) and then explored as moderators of the size of treatment effects as measured using the PND. A consistent finding across meta-analytic reviews is that functional assessment significantly increases the effectiveness of the behavioral treatment of severe behavior problems (Carr et al., 1999; Didden, Duker, & Korzilius, 1997; Scotti, Evans, Meyer, & Walker, 1991; Sternberg, Taylor, & Babkie, 1994).

A difficulty with many studies of self-injury is that they do not report data on generalization and maintenance of treatment effects (Scotti, Evans, Meyer, & Walker, 1991). It is important to ask about the effectiveness of treatment both in general and in terms of their success according to these two dimensions. An effective treatment must generalize to new situations. This is something that can be made more likely by attending to generalization issues (e.g., implementing treatment in multiple contexts). However, planning for generalization could be applied similarly to a treatment based on functional assessment or one based on behavior modification. Perhaps a more crucial issue is actually the long-term maintenance of treatment effects. Studies of long-term maintenance of interventions for self-injury are very rare. A notable exception is Durand and Carr's (1992) research, which showed that Functional Communication Training (FCT) and timeout were initially successful to a similar extent, but that the effects of FCT were more likely to maintain. In terms of generalization, there were also data suggesting that FCT but not timeout also generalized to teachers naïve about the interventions.

Without a large evidence base, we need to return to an understanding of a key principle of behavior to explore whether we would expect behavior modification or a treatment based on functional assessment to be more effective in the longer term. A key criticism of early behavior modification was the potential for the problem of response substitution (Kazdin, 1982). As behavior modification strategies do not address the 'underlying causes' of behavior problems, then simply reducing the behavior may have unpredictable effects with apparently new behavior problems emerging to take their place. This potential problem can be understood with reference to the construct of the response class.

A response class consists of behaviors that serve the same function (i.e., are functionally but not necessarily topographically equiva-

lent). Treatments for self-injury such as FCT (Bird, Dores, Moniz, & Robinson, 1989; Carr & Durand, 1985) make direct use of this concept of functionally equivalent responses. In FCT, the social environment is arranged such that self-injury is not reinforced but a functionally equivalent socially appropriate behavior (e.g., vocal or signed request for attention or help with difficult tasks) is reinforced. The idea is that as the more 'acceptable' member of the response class more reliably leads to reinforcement, self-injury is less likely to occur (Scotti, Evans, Meyer, & DiBenedetto, 1991). Principles related to the matching law (McDowell, 1988; Myerson & Hale, 1984; Pierce & Epling, 1995) would predict such a state of affairs, as organisms will tend to behave in a manner that maximises reinforcement.

Returning to the question of the relative effectiveness of a functional approach versus behavior modification, one would predict that treatment effects achieved through behavior modification would be less likely to maintain. This would be for two main reasons. First, there would be no consistent selection of an appropriate alternative response, and so the emergence of behavior from the same response class to achieve the same function would be unpredictable. The 'new' behavior in these circumstances could well be another self-injury topography or another problem behavior. Second, without an explicitly developed alternative response that is functional in a range of environments (note that this is also a generalization issue), a client would be very likely to revert to self-injury once the behavioral programme in place was ended or once its effective implementation began to deteriorate.

Treatment fidelity, or the extent to which a procedure is implemented accurately, is another level at which we might consider the effectiveness of treatments for self-injury. A final dimension is the likely effect of beginning treatment on frequency of self-injury due to the phenomenon of the extinction burst. When reinforcement is removed, the affected behavior tends to occur at higher frequencies for a brief period. The significance of both treatment fidelity and the phenomenon of the extinction burst in the present discussion can be illustrated by describing a research study conducted by McConnachie and Carr (1997).

McConnachie and Carr (1997) found that interventions for self-injury based on a FCT approach led to a reduced extinction burst when compared to escape extinction. Given that one would wish ethically to minimize the occurrence of self-injury, an intervention based on a functional assessment may be more effective at this level. A further key result of the study was that when staff were not being monitored closely by supervisors, those using escape extinction were less consistent in applying their intervention. McConnachie and Carr suggest that these two phenomena are linked in that staff intervention behavior is maintained by avoidance contingencies. Challenging behaviours such as self-injury are aversive (Mossman, Hastings, & Brown, 2002) and through avoiding them, counter-habilitative staff behaviors can be negatively reinforced (Hastings & Remington, 1994).

Least Restrictive Treatment Procedures

A second ethical principle guiding the work of behavior analysts is that the least restrictive treatment alternative should be used (Johnston & Sherman, 1993). The use of punishment procedures is not explicitly ruled out, although behavior analysts are encouraged to use reinforcement procedures whenever possible. The underlying goal of a behavior modification approach to self-injury is to remove or reduce the frequency of the behavior. This approach might also be called eliminative (Goldiamond, 1974). An approach based on a functional assessment is constructional (Goldiamond), as new behaviours are learned to replace self-injury. Given the focus on learning of new behavior, or development of behaviors already in the client's repertoire, treatment based on a functional assessment is by definition likely to be less restrictive and thus ethically preferable.

There may be an argument, however, for the use of restrictive behavior modification techniques when self-injury is extreme and/or immediately life-threatening. Punishment correctly applied can be effective in rapidly reducing self-injury. For example, Romanczyk and Goren (1975) showed that contingent electric shock could be used to suppress self-injury. Thus, there may be situations in which one could argue that the first goal of treat-

ment is to reduce the risk of self-injury and thus it is ethical to use a restrictive procedure.

Social Validity

A third ethical issue of relevance to the treatment of self-injury is selecting treatments that lead to socially valued outcomes via socially valued means. The BACB guidelines are not particularly clear on this issue, referring in the section on research issues to the expectation that behavior analysts act in a manner that maintains the dignity of participants. Socially valued outcomes are also explicit in the description of ABA as Applied (Baer et al., 1968). The concept of the social validity of treatments has, however, also been an important issue in behavior analysis for some years.

Behavior analysts have espoused the principle that data on acceptability of interventions (to staff, families, wider society, and people with developmental disabilities themselves) should help to guide treatment selection. This criterion has been given the general label of 'social validity', or the views of consumers on the importance, acceptability and effectiveness of interventions (Wolf, 1978). Probably the most extensively studied dimension of social validity is the social significance of behavioral intervention procedures themselves (Wolf), operationalized as the appropriateness, fairness, and reasonableness of procedures applied to ameliorate a behavioral problem (Kazdin, 1981).

Researchers have generally targeted potential mediators of interventions, especially staff in developmental disabilities services, to gather data on social validity. Existing studies have focused on a range of factors that may affect staff views on social validity, but the most reliable effects have proved to be the type of intervention, and severity of the presenting problem (Elliott, 1988; Lennox & Miltenberger, 1990; Miltenberger, 1990; Reimers, Wacker, & Koeppl, 1987; Storey & Horner, 1991). Specifically, reinforcement-based interventions (e.g., differential reinforcement) and interventions that are relatively less intrusive are rated as more acceptable, and the acceptability of intrusive interventions increases as problem severity increases.

A more important question for the present discussion is whether interventions based on functional assessment are perceived as more socially acceptable by consumers. Research studies published to date cannot be used to draw a strong conclusion with one study finding that interventions linked to functional assessment are rated as more socially valid (Jones & Lungaro, 2000), two studies finding weak or inconsistent effects (Hastings, Boulton, Monzani, & Tombs, 2004; Weigle & Scotti, 2000), and one study finding no effects (Miltenberger & Lumley, 1997).

Further Issues

In addition to the main questions reviewed above, there are some further ethical issues that have a bearing on the balance of the debate. Those discussed here relate to: 1. The time and personnel investment in conducting functional assessment, especially analog assessment, 2. The use of preventative interventions, 3. The elicitation of self-injury during analog assessment procedures, and 4. The clarity of results from functional assessment methods. Each of these is discussed in turn below.

Functional assessment takes time and specialist resources to carry out. In the case of a full analog assessment, data collection takes hours over several days and also requires considerable training in observational skills and data presentation and interpretation (see Iwata et al., 1982). Even for non-experimental methods, an experienced behavior analyst must spend considerable time carrying out analyses and the time input from other service staff to provide data for a functional assessment is significant. Although brief methods of analysis have been developed (e.g., Kahng & Iwata, 1999; Northup et al., 1991), these still require an expert team and are often conducted in a specialized outpatient setting. The general question is whether time and resources would be best employed simply teaching clients key skills such as functional communication techniques. Given that around 70% of functional assessments seem to identify either attention or demand avoidance functions for challenging behaviors including self-injury (Derby et al., 1992; Iwata et al., 1994), appropriate behaviors for achieving these two functions could perhaps be taught

during the time that it takes to conduct functional assessments.

This last point also relates to a broader question about preventative intervention. Self-injurious behaviors do not appear from nowhere, but are shaped over time by exposure to environmental and especially social contingencies (Guess & Carr, 1991). This developmental process can be understood partly through the concept of the establishing operation (Michael, 1982). Michael promoted this concept as a way of offering a behavior analytic explanation of motivation. Fundamentally, the motivation for an individual to obtain positive reinforcement occurs as a direct consequence of a period of deprivation of that reinforcement. Similarly, the motivation for negative reinforcement occurs as a direct consequence of exposure to an aversive event. Applying this notion to the development of self-injury, these behaviors may function to overcome a state of deprivation (such as a period of no social contact or activity), or to avoid an aversive situation (such as an academic demand). Without the environmental conditions to establish self-injury as reinforcing, the behavior would not occur.

Thus, getting to the stage of the need for a functional assessment is in many ways too late. Intervening early to identify establishing operations that increase likelihood of self-injury occurring is an obvious therapeutic goal. It may involve preventing states of deprivation (e.g., lack of social attention), through techniques such as non-contingent reinforcement (Vollmer, Iwata, Zarcone, Smith, & Mazaleski, 1993), to enabling more adaptive means of avoiding demands or gaining attention through teaching functional communication skills (Dunlap, Foster-Johnson, & Robbins, 1990; Reeve & Carr, 2000). Ultimately, such preventative strategies may prove to be more efficient and effective (and thus ethical) treatment approaches.

The preceding discussions about intervening either without a functional assessment or earlier in the developmental history of self-injurious behaviors are important points relating to the merits of conducting functional assessments versus not doing so. However, there is a broader issue about these interventions. We characterized behavior modification as the application of behavior change procedures without a functional assessment, and a strict interpretation of this definition might encompass interventions designed to teach skills such as functional communication without an explicit functional assessment. However, the interventions in these cases would arguably be based on an implicit functional assessment built on the basis of evidence across a large number of published cases. There is also another sense in which these general preventative interventions cannot be considered as simply behavior modification and that is the fact that they are clearly constructional (cf. Goldiamond, 1974) in nature.

A further issue is that experimental analog assessment is designed to elicit self-injury under controlled experimental conditions. Given that one would want to avoid unnecessary harm to clients, it is important to consider that self-injury can be seen to be 'made' to occur within analog assessment purely to enable hypotheses about its function to be derived. Thus, behavior analysts would need to be convinced of the validity of analog assessment methods in balancing the ethical case for their use. However, there is a range of internal and external validity problems with the experimental conditions that are a constituent part of analog assessment (Oliver, 1991; Sturmey, 1995). For example, there is a possibility that self-injury could be shaped within the analyses by contingencies that have not in fact been responsible for maintaining the behavior in the natural environment. Therefore, the ethical case for use of analog assessment is not watertight. This particular criticism does not necessarily extend to all functional assessment methods.

A final ethical point to be discussed here is that outcomes of functional assessments are not always clear. In particular, no clear function or multiple functions might be identified in a significant minority of cases (Derby et al., 1992; Iwata et al., 1994). When such results are obtained, the links to intervention are much less clear and it is then questionable whether the effort in conducting a functional assessment has been worthwhile. Of course, there is a balance of risk here in that for many cases of self-injury it is possible to identify clear functional hypotheses and that these may be the optimal conditions for a maximally

effective and thus ethically superior treatment.

Conclusions

The present review suggests that basing treatment of self-injury on results of a prior functional assessment is more ethical than an eliminative behavior modification approach. Firstly, treatment based on functional analysis is more effective overall, more likely to generalize and maintain, and is probably more likely to be implemented with good fidelity by staff (presumably increasing the chance of successful treatment). Secondly, to the extent that treatments based on functional assessments are more likely to be constructional in nature they are also likely to be less restrictive than behavior modification approaches. Thirdly, although direct data from the small number of empirical studies are ambiguous, to the extent that procedures based on functional assessment use reinforcement-based strategies they are likely to be viewed by consumers as more acceptable.

These conclusions lend support to use of functional assessments in the treatment of self-injury. However, it is also important to consider limitations of functional assessment approaches. These include the intensity of resources required, a potential lack of clarity of results, and questions about validity. Furthermore, there may be specific situations where self-injury is immediately life-threatening when a behavior modification procedure is a swift and ethical intervention. These factors and others may limit the ethical superiority of treatments based on functional assessment and indicate issues to be addressed in the further development of functional assessment and functional interventions for self-injurious behavior.

References

BACB. (2001). *Guidelines for responsible conduct for behavior analysts.* Behavior Analyst Certification Board.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Bird, F., Dores, P. A., Moniz, D., & Robinson, J. (1989). Reducing severe aggressive and self-injurious behaviors with functional communication training. *American Journal on Mental Retardation, 94,* 37–48.

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

Carr, E. G., Horner, R. H., Turnbull, A. P., Marquis, J. G., McLaughlin, D. D., McAtee, M. L., et al. (1999). *Positive behavior support for people with developmental disabilities: A research synthesis.* Washington, DC: American Association on Mental Retardation.

Collacott, R. A., Cooper, S. A., Branford, D., & McGrother, C. (1998). Epidemiology of self-injurious behaviour in adults with learning disabilities. *British Journal of Psychiatry, 173,* 428–432.

Derby, K. M., Wacker, D. P., Sasso, G., Steege, M., Northup, J., Cigrand, K., et al. (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.

Didden, R., Duker, P. C., & Korzilius, H. (1997). Meta-analytic study on treatment effectiveness for problem behaviors with individuals who have mental retardation. *American Journal on Mental Retardation, 101,* 387–399.

Dunlap, G., Foster-Johnson, L., & Robbins, F. R. (1990). Preventing serious behavior problems through skill development and early intervention. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 273–286). Sycamore, IL: Sycamore Publishing.

Durand, V. M., & Carr, E. G. (1992). An analysis of maintenance following functional communication training. *Journal of Applied Behavior Analysis, 25,* 777–794.

Elliott, S. N. (1988). Acceptability of behavioral treatments: A review of variables that influence treatment selection. *Professional Psychology: Research and Practice, 19,* 68–80.

Emerson, E., Kiernan, C., Alborz, A., Reeves, D., Mason, H., Swarbrick, R., et al. (2001). The prevalence of challenging behaviors: A total population study. *Research in Developmental Disabilities, 22,* 77–93.

Goldiamond, I. (1974). Toward a constructional approach to social problems: Ethical and constitutional issues raised by applied behavior analysis. *Behaviorism, 2,* 1–84.

Guess, D., & Carr, E. G. (1991). Emergence and maintenance of stereotypy and self-injury. *American Journal on Mental Retardation, 96,* 299–319.

Hastings, R. P. (2002). Do challenging behaviors affect staff psychological well-being?: Issues of

causality and mechanism. *American Journal on Mental Retardation, 107,* 455–467.

Hastings, R. P., Boulton, H. V. N., Monzani, L. C., & Tombs, A. K. H. (2004). Behavioral function effects on intervention acceptability and effectiveness for self-injurious behavior. *Research in Developmental Disabilities, 25,* 139–154.

Hastings, R. P., & Remington, B. (1994). Rules of engagement: Toward an analysis of staff responses to challenging behavior. *Research in Developmental Disabilities, 15,* 279–298.

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

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., et al. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis, 27,* 215–240.

Johnston, J. M., & Sherman, R. A. (1993). Applying the least restrictive alternative principle to treatment decisions: A legal and behavioral analysis. *Behavior Analyst, 16,* 103–115.

Jones, K. M., & Lungaro, C. J. (2000). Teacher acceptability of functional assessment-derived treatments. *Journal of Educational and Psychological Consultation, 11,* 323–332.

Kahng, S., & Iwata, B. A. (1999). Correspondence between outcomes of brief and extended functional analysis. *Journal of Applied Behavior Analysis, 32,* 149–160.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12,* 493–506.

Kazdin, A. E. (1982). Symptom substitution, generalization, and response covariation: Implications for psychotherapy outcome. *Psychological Bulletin, 91,* 349–365.

Lennox, D. B., & Miltenberger, R. G. (1990). On the conceptualization of treatment acceptability. *Education and Training in Mental Retardation, 25,* 211–224.

McConnachie, G., & Carr, E. G. (1997). The effects of child behavior problems on the maintenance of intervention fidelity. *Behavior Modification, 21,* 123–158.

McDowell, J. J. (1988). Matching theory in natural human environments. *The Behavior Analyst, 11,* 95–108.

Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior, 37,* 149–155.

Miltenberger, R. G. (1990). Assessment of treatment acceptability: A review of the literature. *Topics in Early Childhood Special Education, 10,* 24–38.

Miltenberger, R. G., & Lumley, V. A. (1997). Evaluating the influence of problem function on treatment acceptability. *Behavioral Interventions, 12,* 105–111.

Mossman, D. A., Hastings, R. P., & Brown, T. (2002). Mediators' emotional responses to self-injurious behavior: An experimental study. *American Journal on Mental Retardation, 107,* 252–260.

Myerson, J., & Hale, S. (1984). Practical applications of the matching law. *Journal of Applied Behavior Analysis, 17,* 367–380.

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

Oliver, C. (1991). The application of analogue methodology to the functional analysis of challenging behaviour. In B. Remington (Ed.), *The challenge of severe mental handicap: A behavior analytic approach* (pp. 97–118). Chichester, UK: Wiley.

Oliver, C., Murphy, G. H., & Corbett, J. A. (1987). Self-injurious behaviour in people with mental handicap: A total population study. *Journal of Mental Deficiency Research, 31,* 147–162.

Pierce, W. D., & Epling, W. F. (1995). The applied importance of research on the matching law. *Journal of Applied Behavior Analysis, 28,* 237–241.

Reeve, C. E., & Carr, E. G. (2000). Prevention of severe behavior problems in children with developmental disorders. *Journal of Positive Behavior Interventions, 2,* 144–160.

Reimers, T. M., Wacker, D. P., & Koeppl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review, 16,* 212–227.

Repp, A. C., Felce, D., & Barton, L. E. (1988). Basing the treatment of stereotypic and self-injurious behaviors on hypotheses of their causes. *Journal of Applied Behavior Analysis, 21,* 281–289.

Repp, A. C., Singh, N. N., Olinger, E., & Olson, D. R. (1990). The use of functional analysis to test causes of self-injurious behaviour: Rationale, current status, and future directions. *Journal of Mental Deficiency Research, 43,* 95–105.

Romanczyk, R. G., & Goren, E. R. (1975). Severe self-injurious behavior: The problem of clinical control. *Journal of Consulting and Clinical Psychology, 43,* 730–739.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single subject research: Ideas and applications. *Behavior Modification, 22,* 221–242.

Scotti, J. R., Evans, I. M., Meyer, L. H., & DiBenedetto, A. (1991). Individual repertoires as behavioral systems: Implications for program design and evaluation. In B. Remington (Ed.), *The challenge of severe mental handicap: A behavior ana-*

lytic approach (pp. 139–163). Chichester, UK: Wiley.

Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behaviors: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96,* 233–256.

Sternberg, L., Taylor, R. L., & Babkie, A. (1994). Correlates of interventions with self-injurious behaviour. *Journal of Intellectual Disability Research, 38,* 475–485.

Storey, K., & Horner, R. H. (1991). An evaluative review of social validation research involving persons with handicaps. *Journal of Special Education, 25,* 352–401.

Sturmey, P. (1995). Analog baselines: A critical review of the methodology. *Research in Developmental Disabilities, 16,* 269–284.

Vollmer, T. R., Iwata, B. A., Zarcone, J. R., Smith, R. G., & Mazaleski, J. L. (1993). The role of attention in the treatment of attention maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. *Journal of Applied Behavior Analysis, 26,* 9–21.

Weigle, K. L., & Scotti, J. R. (2000). Effects of functional analysis information on ratings of intervention effectiveness and acceptability. *Journal of the Association for Persons with Severe Handicaps, 25,* 217–228.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203–214.

Received: 21 April 2004
Initial Acceptance: 30 June 2004
Final Acceptance: 1 December 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 343–351

Review of Assessment Procedures for Students with Moderate and Severe Disabilities

Ellin Siegel
University of Nebraska-Lincoln

Rose M. Allinder
Rafiki Children's Center, Nairobi, Kenya

Abstract: Assessment procedures for students with moderate or severe disabilities cover a range of procedures and practices. This article reviews recommendations for assessment practices for these students. These recommendations are compared with the assessment actually conducted by a sample of 22 teachers of students with moderate and severe disabilities. Three assessment purposes were analyzed; diagnosis and placement, curriculum and program development, and evaluation. Results are discussed in terms of best practices and recommendations for assessment of students with moderate or severe disabilities.

Assessment procedures for students with moderate and severe disabilities represent a conundrum. Students must be assessed to be classified and placed into special education, yet these assessments rarely yield information useful to practitioners for designing or evaluating effects of programming (Snell & Brown, 2006). Sigafoos, Cole, and McQuarter (1987) noted this dilemma in their investigation of then-current assessment practices of students with severe disabilities. Their analysis of assessment practices indicated that over 50 percent of students with severe disabilities had been administered standardized, norm-referenced tests (NRTs), which evidenced varying degrees of reliability and validity. In contrast, use of criterion-referenced tests and adaptive behavior scales was much less common. As Sigafoos et al. point out, use of NRTs, while required for classification purposes, requires supplementation by other sources of data. This is necessary to ensure that data generated by the assessment process would be "relevant for planning and tracking on going progress" (Wolf-Schein, 1998, p. 52-53).

The necessity of additional types of assessment becomes more apparent when the purposes of assessment are more than classification and identification. Two critical areas that require assessment procedures that are carefully chosen or created, administered and interpreted are program development and instructional evaluation (Grisham-Brown, 2000; Macfarlane, 1998; Siegel-Causey & Allinder, 1998; Snell & Brown, 2006). For each of these, a range of assessment practices is promoted to meet best-practice standards.

Program development, in which an individual education plan (IEP) for the student is developed, *may* be grounded in the assessment done for classification and placement purposes (Alper, Ryndak, & Schloss, 2001). Assessment for these purposes includes, at minimum, evaluation of intelligence and adaptive behavior. If data drawn from assessment for eligibility purposes are used to develop IEPs, several caveats must be issued. One relates to use of intelligence tests for these students. The appropriateness of these tests has been questioned for a number of reasons. First, for students who have accompanying disabilities, which make responding in standard format impossible, modification of response type makes validity and reliability unknown. Second, validity and reliability is

This study was conducted as part of Project CAAR: Creating Alternative Assessment Results. We extend our thanks to the teachers in Lincoln Public Schools, Special Education Division who participated in our portfolio research for students with moderate or severe disabilities. We also thank Julie Ortgies for her role as research assistant with the project. Correspondence concerning this article should be addressed to Ellin B. Siegel, Department of Special Education and Communication Disorders, Communication 318 Barkley Memorial Center, University of Nebraska-Lincoln, Lincoln, NE 68583.

questionable unless tests' normative populations include students with the same range and level of cognitive functioning (Salvia & Ysseldyke, 2001). A final concern with using intellectual functioning for program development purposes, for students who are able to complete them, is that these tests were designed to compare them with their age peers, not to document what students do and do not know or can and cannot do.

In addition, caveats for using adaptive behavior assessment results must be noted. While this type of assessment also yields a norm-referenced score similar to that obtained in intelligence tests, it is possible to use adaptive behavior assessment for program development purposes by examining ratings of specific items within the domains of the adaptive behavior scale. Using adaptive behavior scales for this purpose, however, must consider if they are conducted only tri-annually. The information may be too out-dated to be of use for annual IEP evaluation and planning. Practitioners are urged to re-assess adaptive behavior more often when updating and revising IEPs.

While adaptive behavior scales may be used for program development purposes, other types of assessments should be used to develop each student's IEP. These other assessments may include criterion-referenced tests (CRTs) and ecological inventories (Snell & Brown, 2006). CRTs measure the degree to which a student has obtained specific skills. As Beirne-Smith, Ittenbach, and Patton (2002) point out, CRTs vary with regard to what skills are assessed and the specificity with which skills are assessed. Practitioners are urged to use care in selecting CRTs to use with students with more severe disabilities and may find that creating their own CRTs is more beneficial for the student. Ecological inventories are a form of assessment in which activities and skills necessary for different environments, such as community recreation, vocational, domestic or general education settings are analyzed (Browder, 2001; Downing, 2002). A student's performance on the ecological inventory contributes to the IEP both for the present level of performance and goal-setting sections.

Some students with moderate disabilities may require assessment of functional academic skills. CRTs and NRTs may provide insight into the degree of academic skill development as well as pinpointing specific functional academic skills to be targeted for instruction.

Using either CRTs or ecological inventories or some other method of assessing students for the purpose of educational programming can result in an overwhelming number of concerns. In this mix of data, family concerns, interests, and desires may be slighted. This should be avoided because creating education programs around family concerns, interests and desires may increase the likelihood that the student will obtain and maintain educational goals. The importance of building students' programs around family concerns is underscored by IDEA 97 regulations that require that "concerns of the parents for enhancing the education of their child" (34 CFR 300.346) be incorporated into the IEP process. Several processes have been developed which help uncover families' visions and hopes and dreams for their children and which help delineate important and meaningful goals for individual students. These processes are intentionally strengths-based, focusing not merely on the needs and deficits of the student with moderate or severe disabilities, but their strengths. These methods include personal futures (Mount & Zwerkik, 1988), Making Action Planning System (MAPS; Forest & Snow, 1987), and Choosing Outcomes and Accommodations for Children (COACH; Giangreco, Cloninger, & Iverson, 1998).

In personal futures or lifestyle planning, concerned family members, friends, advocates and professionals meet to collaboratively evaluate an individual's quality of life and describe a desirable future for that person (Browder, 2001). MAPS (Forest & Snow, 1987) is a collaborative team approach to identifying critical information about the student with severe disabilities. Family members, professionals, advocates, and peers relate their dreams, nightmares, and descriptions of the student as well as their descriptions of the student's strengths, needs, and ideal day. Responses to these questions can aid in developing and prioritizing goals for the student's IEP (Forest & Lusthaus, 1990). Similarly, COACH (Giangreco et al., 1998) is used by school personnel and family members to identify IEP goals and objectives in accordance with family-centered

priorities, using a structured family-interview approach.

After the educational program has been developed and instruction begins, it is critical to assess student progress. As Wolf-Schein (1998) points out, "educational assessment is useful insofar as it contributes to positive development of the child." (p. 38) The best care in selecting, administering, and interpreting eligibility assessment information or program development information is pointless if the child does not make progress. Typical ways of assessing progress in other students may be of limited use for students with moderate or severe disabilities. It is important that assessments, which are sensitive to small changes, are used. This may be various ongoing assessments such as informal anecdotal recording keeping, task analytic assessment, discrete trial or opportunity assessment, curriculum-based measurement, or time-based assessment (Farlow & Snell, 1994; Macfarlane, 1998).

While expert advice is to craft assessment practices in alignment with best-practices standards (e.g., Macfarlane, 1998; Siegel-Causey & Allinder, 1998) it is not known to what degree practitioners implement these recommendations. Sigafoos et al. (1987) studied assessment practices for a large number of students with severe disabilities by examining students' cumulative files. Results of that study indicated that tests used with these students were lacking in technical qualities such as reliability and validity and that measures other than NRTs such as CRTs and adaptive behavior measures were used less frequently than recommended. This study did not investigate informal assessment practices to collect student performance data in order to make educational decisions such as programming development and instructional evaluation. The current study was designed to do this. Specifically, we examined assessment practices of 22 teachers of students with moderate or severe disabilities across the different purposes of assessment.

Method

Participants

Participants were 22 special education teachers from different schools in an urban school district in the Midwest. Teachers were chosen from volunteers who had students with moderate or severe disabilities and who had not been trained previously to use a specific form of alternative assessment with their students. Both teachers and parent(s) of participating students completed consent forms.

All but one of the teachers were employed full-time. Median age of teachers was between 40–49 and; the mean number of years teaching experience was 12.1 (range 15-23). Thirteen had a bachelor's degree and nine had a master's degree.

Participating students were in Kindergarten through fourth grade with a mean age of 9 years 2 months (range 5.0–11.2). Each child was classified using state eligibility criteria as mentally handicapped; moderate or severe/profound; multi-handicapped; or autistic. Table 1 details student demographics.

Procedure

We were interested in learning about all assessments used with students in this study. Therefore, we examined student records from cumulative files and from teachers' records, limiting assessment information to the previous five years. Information obtained was photocopied at the student's school and student names were blacked out immediately and replaced with a code number. To obtain this information, teachers were asked to make available to project staff all assessment information for the target student. To verify that project staff obtained copies of *all* assessments for each child, two forms of follow up were conducted. First, project staff used a previously-generated list of possible assessment information and interviewed teachers at their schools to see if some items on the list were available for their student but had not been given to the project staff. Second, teachers were given a listing of possible assessments, which indicated what had been obtained by project staff for their student. Teachers were asked to review the listing and to contact staff if there were errors (e.g., assessments existed but had not been given to the staff).

Once all assessment information was obtained, we categorized each item into one of three categories: (a) diagnosis and placement, (b) program development, and (c) instruc-

TABLE 1

Student demographics

Student	*Gender*	*Age*	*Grade*	*Primary disability label*
A. D.	Male	11.2	5	Mental handicap: severe profound
B. L.	Male	8.0	2	Multiple disabilities
C. T.	Male	9.10	3	Other health impairment
C. T.	Male	11.0	4	Other health impairment
C. N.	Female	6.5	1	Multiple disabilities
E. E.	Male	7.6	2	Autism
E. T.	Male	8.2	3	Autism
E. T.	Male	9.9	3	Autism
H. Y.	Male	10.11	4	Mental handicap: severe profound
H. E.	Female	12.0	5	Mental handicap: moderate
H. S.	Female	7.9	K-1	Mental handicap: moderate
H. N.	Female	8.10	2	Orthopedically impaired
H. N.	Female	10.11	4	Mental handicap: moderate
K. A.	Female	5.0	K	Autism
M. Y.	Male	10.3	4	Mental handicap: severe profound
P. G.	Male	10.8	4	Mental handicap: moderate
P. T.	Male	10.6	4	Mental handicap: moderate
P. E.	Male	8.4	3	Autism
R. H.	Female	8.4	3	Multiple disabilities
R. Y.	Male	10.2	5	Mental handicap: moderate
S. N.	Female	10.3	5	Autism
Z. N.	Male	6.7	1	Mental handicap: moderate

tional evaluation (Grisham-Brown, 2000; Macfarlane, 1998; Siegel & Allinder, 1998; Snell & Brown, 2006).

According to Snell and Brown (2006), the process of diagnosis and placement is used to determine the type of disorder displayed by the child and to make decisions regarding eligibility, classification, and placement. Types of assessments that fall under this category include intelligence tests, adaptive behavior measures, and motor and sensory functioning measures (Taylor, 1997). These assessments usually are administered by a psychologist or specialist (e.g., physical therapist, vision specialist). Program development, the second category, refers to assessments that are used to ascertain skills that should be taught and are used in the development of the child's IEP. This type of assessment is frequently conducted by the teacher in the classroom environment. Finally, assessments are used to evaluate the educational program of the child, including student progress. Assessments used here may include taking frequent probe data, curriculum-based measurement, or frequent and informal anecdotal recording keeping (Farlow & Snell, 1994). We compared types of assessment recommended by best practice literature for each assessment purpose with what this sample of teachers actually did.

Results

Diagnosis and Placement

By state statute, students identified as having a mental handicap must be administered a cognitive and an adaptive behavior measure. In our sample of students, seven participating students had an intellectual assessment. For seven other students, the multi-disciplinary team (MDT) had determined that formal cognitive assessment was not possible due to the severity disability. Specific tests administered for cognitive functioning varied and included: the *Kaufman Assessment Battery for Children* (Kaufman & Kaufman, 1983) ($n = 3$); *Bayley Scales of Infant Development* (Bayley, 1984) (attempted with four children, but successfully completed for only two); *Stanford-Binet Intelli-*

gence Scale (Thorndike, Hagen, & Sattler, 1985) (attempted with two children but successfully completely for only one), *Woodcock Johnson Test of Cognitive Ability* (Woodcock & Mather, 1989) (unsuccessfully attempted with one child), and *Cognitive Abilities Test* (Thorndike & Hagen, 1994) ($n = 1$).

In the domain of adaptive behavior, all but one student received some type of assessment and several students received more than one. The *Vineland Adaptive Behavior Scale* (Sparrow, Balla, & Cicchetti, 1984) was the most common ($n = 18$). Other measures were: *Scales of Independent Behavior* (Bruininks, Woodcock, Weatherman, & Hill, 1984) ($n = 3$), *Maxfield-Buchholz Social Maturity Scale for Blind Preschool Children* (Maxfield & Buchholz, 1957) ($n = 2$) and *Developmental Assessment of Life Experiences* (Barber et al., 1995) ($n = 3$).

In addition to assessing cognitive and adaptive behavior skills, most students had other NRTs administered. In the domain of speech/ language, all but one had some assessment. Informal assessment, including observation of the child and/ or interview of the child's parent(s) and/ or teacher, was common ($n = 20$). Among formal assessments administered, the *Peabody Picture Vocabulary Test- Revised* (Dunn, 1984) was the most prevalent ($n = 6$), followed by the *Arizona Articulation Proficiency Scale* (Fudala, 1987) ($n = 4$), the *Preschool Language Scale - 3* (Zimmerman, Steiner, & Pond, 1979) ($n = 4$), the *Nonspeech Test* (Huer, 1988) ($n = 4$), the *Expressive One-Word Picture Vocabulary Test* (Gardner, 1979) ($n = 3$), the *Rossetti Infant-Toddler Language Scale* (Rossetti, 1990) ($n = 2$), Communication and *Symbolic Behavior Scale* (Wetherby & Prizant, 1993) ($n = 2$), and four others.

Only five students were administered formal tests of academic ability, although one subject was unable to complete the test. Assessments used included *Woodcock-Johnson Psycho-Educational Battery, Tests of Achievement* (Woodcock & Johnson, 1989) ($n = 2$) and the *Psycho-Educational Profile- Revised* (Schopler, Rechler, Bashford, Lansing & Marcus, 1990) ($n = 2$).

As part of their diagnosis and placement assessment procedures, 15 students were assessed in the domain of behavior. The majority of these ($n = 11$) were by means of informal observation. The remaining assessments administered were varied and included the *Social Skills Rating System* (Gresham & Elliott, 1990) ($n = 2$) and the *Child Behavior Checklist* (Achenbach, 1991) ($n = 2$) among others.

The *Childhood Autism Rating Scale* (Schopler, Reichler, & Renner, 1988) was administered to a small number of children ($n = 4$). Of the 22 students in the sample, six had been diagnosed with autism. Because of the ongoing nature of this disorder, it is likely that the other two children had been diagnosed prior to the time limits of this study and a re-evaluation may have been considered unnecessary.

Motor skills were assessed by an occupational therapist or a physical therapist for 15 students. In addition, vision and hearing tests were administered to 13 and 15 students respectively. About half of those tested for vision and hearing, however, were unable to complete the test successfully.

Curriculum and Program Development

Best practice standards recommend that information to develop the individual's instructional program be drawn from a variety of sources, including adaptive behavior measures, CRTs of academic skills, ecological inventories and parent input. Among students in our sample, only five had any documentation of caregiver input (such as an interview or survey) for planning the IEP, although the fundamental instrument in developing the curriculum is the IEP.

Several teachers administered CRTs to gain additional information. For example, the *Generic Skills Inventory* (GSI) (Oregon Department of Education, 2001), which has subcategories of motor, cognition, dyadic/social, receptive and expressive communication, and self-help, was administered to four students. Other CRTs included: *Brigance Diagnostic Inventory of Early Development* (Brigance, 1978), completed for one student; *MacArthur Communicative Development Inventories* (Fenson et al., 1993) completed for one student; and various language/communication assessments. Academically, teachers completed the school district criterion-referenced measures in reading and math for eight students and five students, respectively. These measures were multi-year synopses of students' progress through devel-

opmental skill levels of reading and math, which were completed and maintained for general education students.

Augmentative communication objectives and recommendations were used in the curriculum development for some students. In addition, for students who changed schools, a Transition Planning form or checklist was used to facilitate transition to the new school environment.

As an alternative to formal testing, individualized assessments may provide information that is more useful for a given child with special needs (Snell & Brown, 2006). Nearly one-third of the teachers created some form of a task or functional analysis, specific to their student's needs. For example, one teacher used a functional analysis to define specific steps involved for a student who makes purchases at a grocery store in the local community.

Specialists often assisted to develop or improve programs for students in this sample. These specialists were school district personnel with expertise in programming for students with more severe disabilities (referred to as inclusion cadre personnel) or in behavioral interventions (behavioral specialists), and who consulted with individual teachers. Inclusion cadre was found for nearly three fourths of the students ($n = 16$). Additionally, behavior specialists and other specialists were involved in the program development of some children. Approximately one-third of the students ($n = 7$) attended clinics outside the school setting. Reports from these clinics suggested specific recommendations for the given student.

Evaluation

Not surprisingly, all students had documentation in the form of IEP progress reports or student progress reports (regular or narrative), which are both evaluative documentation required by the local school district.

Some teachers ($n = 8$) maintained a chart that listed IEP goals and objectives and used it to document ongoing progress. Other methods used by teachers to track growth included keeping a notebook of anecdotal records ($n = 3$) and keeping samples of student work to indicate progress ($n = 6$). Some teachers ($n = 4$) incorporated a home note system in order facilitate communication with the child's caregiver(s). Additionally, several teachers designed charts to show improvement for specific skills such as behavior ($n = 4$), bathroom activity ($n = 3$), and classroom activities ($n = 2$). Spelling lists ($n = 2$) and high frequency word lists ($n = 2$) were used for a few students to show ongoing development in these specific literacy skills.

Besides the child's homeroom or resource teacher, other individuals were involved in the evaluative process for some children. For example, four students had narrative reports by other teachers (art, PE, etc.). In addition, 14 students had progress reports by an occupational or physical therapist, seven students had reports by a speech/language therapist, and one student had a report regarding neurodevelopment. Finally, two students had evaluative reports for summer school.

Discussion

In our sample of students with moderate and severe disabilities, we compared the actual assessment processes with those recommended by best practice standards. This comparison yielded a mixed review of how these practices were implemented.

With regard to diagnosis/placement purposes of assessment, there was evidence that few students had been administered a formal norm-referenced test of intelligence. While this measure is required by state statute and definition of mental retardation, it appeared the MDTs demonstrated good clinical judgment in not administering the test to students whose functioning level was severely impaired or to students whose multiple disabilities made it difficult or problematic to ensure that results were valid. Problems associated with administering intelligence tests to students such as were in our sample is not uncommon.

An adaptive behavior measure was administered to 21 students. This is in contrast to results of Sigafoos et al. (1987) that the majority of their sample were not administered this type of assessment. However, we echo the caveat of Sigafoos et al. in stating that we cannot determine how the data from these

measures were used. It is possible that results for adaptive behavior measures were used to fulfill eligibility guidelines and not to plan educational programs.

In addition to these required measures, most students in our sample had additional assessment as part of the referral/placement process. The primary type was for communication, via tests for speech, receptive language, and expressive language.

With regard to assessment for curriculum planning, teachers appeared to follow best practice standards to a lesser degree. Five types of criterion-referenced tests were used with the students in our sample. Additionally, five teachers had conducted task analyses of IEP objectives and one teacher had completed an ecological inventory. Limited evidence of a broader range of CRTs, task analyses, and ecological inventories was disheartening and could lead to the conclusion that students' IEPs might not have been as individualized, comprehensive, or specific as they might have been.

Only one participant had completed a procedure to ascertain family preferences, dreams, desires with regard to their child's future, by completing MAPS. Five other teachers had used the school district one-page form querying families about what goal areas would they like included on their child's IEP. The lack of broad based personal-planning procedures among our participants was discouraging given the emphasis placed on this in the professional literature in recent years and given that the participating school district had previously conducted inservices on this topic.

Regarding assessment practices for program evaluation purposes, the pattern of mixed results continued. Best practices for this assessment purpose include anecdotal recording keeping, task analytic assessment, discreet trial or opportunity assessment, curriculum-based assessment, or time-based assessment. In our sample, we found evidence that all of there were used to some degree. Anecdotal recording was common, both by teachers and by related services personnel such as physical therapists or behavior specialists. Furthermore, there was evidence that a small number of teachers used each of the other types of recommended assessments.

Conclusion

We examined assessment practices and procedures of 22 teachers of students with moderate or severe disabilities. From this examination we found a pattern of mixed evidence that these practitioners were following best practice recommendations. The one area in which these assessment practices appeared to be in closest alignment with best practice recommendations was that of diagnosis/placement. That 21 of 22 students had an adaptive behavior measure completed was encouraging. This is tempered by the supposition that these measures may have been conducted only to fulfill state statute and that results did not appear to have been used when designing students' programs. The area of diagnosis/placement seemed to typify best practice standards for a second reason. Only two students had scores from tests that were normed for a younger age population. This is certainly an improvement over results reported a decade ago (Sigafoos et al., 1987). Additionally, not all students had intelligence scores reported. Given the extent and nature of the disabilities present in our sample of students, this appeared to have been good judgment of case of decision makers.

In contrast to the more positive findings in the diagnosis/placement area, assessments conducted for program planning and program evaluation were less encouraging. Most disappointing was the fact that only one student had results of MAPs reported. Absence of information of this type was perplexing. It is not possible to determine if this was due to lack of training of teachers, recency of students placed on these teachers' caseloads, or difficulties because of more logistical reasons such as lack of time to conduct assessments.

Assessment for program evaluation, however, appears to be even more mixed. Not surprising, anecdotal recording was the most common among our sample of teachers. Much less common were assessments that were tied directly to IEP objectives and which would yield data informing instructional decision making. This was especially problematic from our viewpoint, given research stating that students with more severe disabilities make progress less easily and thus need their growth monitored more closely.

Recommendations

Teachers of students with moderate or severe disabilities should continue to articulate the need for manageable and educationally informative assessment, especially for program planning and evaluation purposes. Moreover, administrators who supervise and support these teachers should be well versed in appropriate assessment procedures and should be able to teach, model, and refine assessment procedures for these students. While assessment is one of the most important topics in education today, assessment encompasses well more than high-stakes assessment. Assessment used for planning individual education programs and evaluating effectiveness of instruction must be emphasized as well.

References

Achenbach, T. M. (1991). *Child Behavior Checklist*-.Burlington, VT: University of Vermont Department of Psychiatry.

Alper, S., Ryndak, D. L., & Schloss, D. N. (2001). *Alternative assessment of students with disabilities in inclusive settings.* Boston: Allyn & Bacon.

Barber, G. A., Olewnik, V., Haughey, J., Gardner, T., Sosnowski, M., Lawrence, J., et al. (1995). *Developmental Assessment of Life Experiences.* Erie, PA: The Barber Center Press, Inc.

Bayley, N. (1984). *Bayley Scales of Infant Development.* San Antonio, TX: Psychological Corp.

Beirne-Smith, M., Ittenbach, R. F., & Patton, J. R. (2002). *Mental Retardation* (6th ed.). Upper Saddle River, NJ: Merrill.

Brigance, A. H. (1978). *Brigance Diagnostic Inventory of Early Development.* North Billerica, MA: Curriculum Associates.

Browder, D. M. (2001). *Curriculum and assessment for students with moderate and severe disabilities.* New York: Guilford Press.

Bruininks, R. H., Woodcock, A., Weatherman, R., & Hill, B. (1984). *Scales of Independent Behavior.* Allen TX: DLM Teaching Resources.

Downing, J. E. (2002). *Including students with severe and multiple disabilities in typical classrooms.* Baltimore: Brookes.

Dunn, L. M. (1984). *Peabody Picture Vocabulary Test - Revised.* Circle Pines, MN: American Guidance Service.

Farlow, L. J., & Snell, M. E. (1995). *Making the most of student performance data.* Washington, DC: American Association on Mental Retardation.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *MacArthur Communicative Development Inventories.* San Diego, CA: Singular.

Forest, M., & Lusthaus, E. (1990). Everyone belongs: With the MAPS action planning system. *Teaching Exceptional Children, 32,* 32–35.

Forest, M., & Snow, J. (1987). *More educational integration.* Toronto: The G. Allan Rocher Institute.

Fudala, J. B. (1987). *Arizona Articulation Proficiency Scale.* Los Angeles: Western Psychological.

Gardner, M. F. (1979). *The Expressive One-Word Picture Vocabulary Test.* Novato, CA: Academic Therapy.

Giangreco, M. F., Cloninger, C. J., & Iverson, V. S. (1998). *Choosing outcomes and accommodations for children: A guide to planning inclusive education.* Baltimore: Brookes.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System.* Circle Pines, MN: American Guidance System.

Grisham-Brown, J. (2000). Transdisciplinary activity-based assessment for young children with multiple disabilities: A program planning approach. *Young Exceptional Children, 3,* 3–10.

Huer, M. B. (1988). *The Nonspeech Test for Receptive/Expressive Language.* Volvo, IL: Don Johnston Developmental Equipment, Inc.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Macfarlane, C. A. (1998). Assessment: The key to appropriate curriculum and instruction. In A. Hilton & A. Ringlaben (Eds.), *Best and promising practices in developmental disabilities* (pp. 35–60). Austin, TX: Pro Ed.

Maxfield, K. E., & Buchholz, S. (1957). *Maxfield-Buchholz Scale of Social Maturity for Use with Preschool Blind Children.* New York: American Foundation for the Blind, Inc.

Mount, B., & Zwerkik, K. (1988). *It's never too early, it's never too late: A booklet about personal futures planning.* St. Paul: Minnesota Governor's Planning Council on Developmental Disabilities. Oregon Department of Education. (2001). *Generic Skills Inventory* (GSI). Salem, OR: Student learning and partnership, Oregon Department of Education.

Rossetti, L. (1990). *Rossetti Infant Toddler Language Scale.* East Moline, IL: Linguisystems.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed.). Boston: Houghton Mifflin Company.

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990) *Psychoeducational Profile Revised* (PEP-R). Austin: PRO-ED

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale.* Los Angeles: Western Psychological Services.

Siegel-Causey, E., & Allinder, R. M. (1998). Using alternative assessment for students with severe

disabilities: Alignment with best practices. *Education and Training in Mental Retardation and Developmental Disabilities, 33,* 168–178.

Sigafoos, J., Cole, D., & McQuarter, J. (1987). Current practices in the assessment of students with severe handicaps. *Journal of the Association of Persons with Severe Handicaps, 12,* 264–273.

Snell, M. E. & Brown, F. (2006). *Instruction of students with severe disabilities* (6th ed.). Upper Saddle River, NJ: Merrill.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Taylor, A. L. (1997). *Assessment of exceptional students: Educational and psychological procedures* (4th ed.). Boston: Allyn & Bacon.

Thorndike, R. L., & Hagen, E. (1994). *Cognitive Abilities Test* (2nd ed.). Chicago: Riverside Publishing Company.

Thorndike, R. L, Hagen, E., & Sattler, J. (1985). *Stanford-Binet Intelligence Scale.* Chicago: Riverside.

Wetherby, A. M., & Prizant, B. M. (1993). *Communication and Symbolic Behavior Scale.* Baltimore: Brookes.

Wolf-Schein, E. G. (1998). Considerations in assessment of children with severe disabilities including deaf-blindness and autism. *International Journal of Disability, Development, and Education, 45*(1), 35–55.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson PsychoEducational Battery-Revised.* Chicago: Riverside.

Woodcock, R. W., & Mather, N. (1989). *Woodcock-Johnson Tests of Cognitive Ability, Standard and Supplemental Batteries.* Allen, TX: DLM.

Zimmerman, I., Steiner, V., & Pond, A. (1979). *Preschool Language Scale - 3.* San Antonio, TX: Psychological Corp.

Received: 3 February 2004
Initial Acceptance: 31 March 2004
Final Acceptance: 25 August 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 352–359

Attitudes of Japanese Adults Toward Persons with Intellectual Disability: An Exploratory Analysis of Respondents' Experiences and Opinions

Toshiaki Tachibana
Aichi Shinshiro Otani University

Abstract: Respondents' experiences most frequently reported in open-ended descriptions of their forming perceptions of persons with intellectual disability were 'passive' experiences, such as coming across them as a stranger in one's town. These were generally not associated with positive attitudes. Formation of positive attitudes mainly derived from experiences such as having a person with intellectual disability in the family, or knowing someone who has a family member with intellectual disability. Another experience associated with positive attitudes was having awareness that anyone could have a person with intellectual disability in his/her family. An unpleasant experience with a person with intellectual disability, especially in childhood, was strongly associated with negative attitudes. Interestingly, most respondents (even those who were in the bottom quartile of attitude scores) believed that they had a more tolerant attitude to disability than the average person.

Perceptions of intellectual disability largely determine attitudes (Antonak, Mulick, Kobe, & Fielder, 1995; Caruso & Hodapp, 1988; Gottlieb & Siperstein, 1976; Tachibana & Watanabe, 2003; Williams, 1986). What factors shape these perceptions? There are few attempts to address this in previous studies. The main purpose of the present study is to throw light on this question. In our previous studies of attitude, we analyzed data based mainly on Likert-type questions (Tachibana & Watanabe, 2004a). Although such data can serve a 'confirmatory' purpose, those alone do not necessarily yield a full picture on the issues under research. At the end of the questionnaire in our study, respondents were required to describe open-endedly their experience of and opinions on people with intellectual disability. This provides information supplementary to that obtained through the confirmatory approach. Thus, our objective is to investigate the relationship between actual experience with persons with intellectual disability, perceptions, and attitude towards such persons on the open-ended responses indications. This will be done through a kind of 'exploratory' analysis because there are no initial hypotheses to be tested.

We express our appreciation to the Municipal Committee for Education in Kasugai and the 11 participant schools for giving us the opportunity for the present survey. We also thank Professor Martin McHugh, National University of Ireland, Galway, for his critical reading of an earlier version of this manuscript. Correspondence concerning this article should be addressed to T. Tachibana, Aichi Shinshiro Otani University, Kawaji, Shinshiro, Aichi 441-1306, JAPAN. E-mail: roku@shinshiro-otani.ac.jp

Method

Participants

Participants were parents (or guardians) of the pupils in 11 elementary schools in Kasugai, Aichi Prefecture, Japan. The 11 schools were selected randomly from all schools (n = 37) in Kasugai. A questionnaire was distributed by teachers to all families (n = 2758) whose children were in attendance. Responses were collected in December, 2000. There were 2381 respondents (females = 2151), males = 230). Mean (*SD*) respondent age was 38.2 (5.0) with a range of 23-65 years.

Questionnaire

The questionnaire had five main sections. In the first, 16 Likert-type questions on attitudes

toward persons with intellectual disability were presented. The second section required respondents to guess the general prevalence of persons with intellectual disability and its future incidence in their own family. They also had to estimate the percentage role of heredity in intellectual disability. In the third section, respondents' perceptions of persons with intellectual disability were elicited. In the fourth section, background variables such as gender, age, occupation and so on were covered. In the last section, data that are analyzed in the present study, the questions were as follows:

I. Please describe open-ended what kind of experience has formed your image of a person with intellectual disability?

II.
   A. Do you think that you have a greater concern than the average person for the problems that people with intellectual disability have? (yes, no)
   B. Please describe open-ended why you have (or have not) a greater concern for these problems than the average person has.

III.
   A. Do you think that you have a more tolerant opinion or a less tolerant opinion than average towards people with an intellectual disability? (more tolerant, less tolerant)
   B. Please describe open-ended why you have a more tolerant (or less tolerant) opinion than average and describe any experience that has formed the opinion, if you have.

(For detailed information on the questionnaire employed, see the previous study, Tachibana & Watanabe, 2004a).

Scoring and Analysis

Categorization of responses. Respondents were grouped on the basis of a frequency analysis of their open-ended replies.

Condensed score. To process the voluminous data, a cluster analysis was performed. This yielded four groups of question items and then four 'condensed scores' of attitudes. We will focus mainly on one condensed score, 'anti-social norm' in the present study. The condensed attitude score of 'anti-social norm' was the mean of scores obtained by Likert-type questions for the following seven attitudes: 1) a hereditary threat to society; 2) marriage with a person who has a family member with intellectual disability; 3) a facility for people with intellectual disability in the neighborhood; 4) living next door to people with intellectual disability in the same apartment building; 5) becoming involved with persons with intellectual disability; 6) making respondent's child sit next to a child with intellectual disability in school; 7) working with people with intellectual disability. (See Appendix in the previous study for these questions, Tachibana & Watanabe, 2004a.) A larger condensed score indicates a more positive attitude toward persons with intellectual disability (Tachibana & Watanabe, 2004a, for full details of how the condensed score were calculated). Also see distribution results of each question (items 1-16) and the condensed score in the previous studies (Tachibana & Watanabe, 2004b, in press).

Statistical analysis. Since the present data were not obtained by random sampling, it is not appropriate to try to infer population values or to calculate p-values (Tachibana, 1988). Instead we will express measures in terms of mean and SE.

Results and Discussion

Formation of perceptions of intellectual disability and attitudes. About 58% of the respondents described some of the experiences that formed their ideas of persons with intellectual disability. Main reported experiences were 13 in number (see Figure 1): 1) having a person with intellectual disability in the family (= family), 2) having a person with intellectual disability amongst one's relatives (= relatives), 3) having a friend who had a child with intellectual disability (= friend), 4) having experience of volunteer work with persons with intellectual disability (= volunteer), 5) having a job which involved service for persons with intellectual disability (= job), 6) having experience of visiting a facility for persons with intellectual disability (= visit), 7) having an acquaintance who had a child with intellectual disability (= acquaintance), 8)

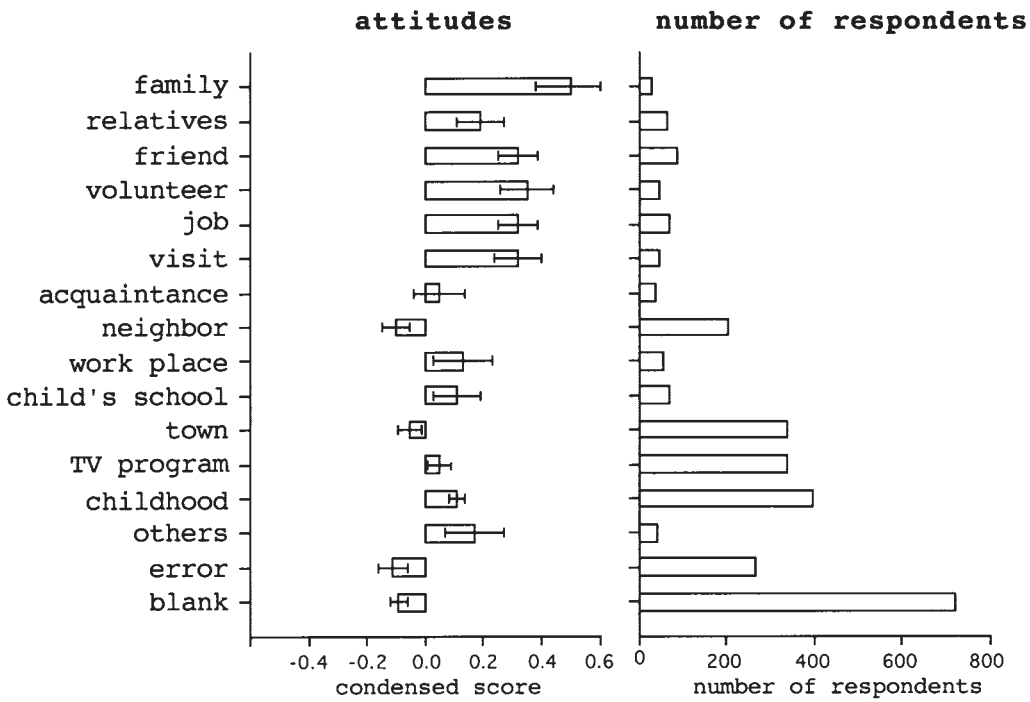

Figure 1. Relation between formative experiences and the condensed attitudes scores toward persons with intellectual disability; a histogram also, giving the number of respondents citing each kind of experience (See the text for explanation of abbreviated terms). A larger condensed score indicates a more positive attitude toward persons with intellectual disability.

having come across a person with intellectual disability in one's neighborhood or in the course of commuting (= neighbor), 9) having a person with intellectual disability as a colleague at work (= work place), 10) having a child with intellectual disability in the same school of one's child (= child's school), 11) having come across a stranger with intellectual disability in one's town (= town), 12) having seen a relevant TV program (= TV program), and 13) having encountered children with intellectual disability in one's childhood (= childhood). About 11% of respondents made the mistake of describing only their image of a person with intellectual disability instead of their experience from which the image was derived (= error). About 30% of respondents did not offer any description (= blank). Seventy-four percent of respondents described just one formative experience, 22% of respondents described two, 3% of respondents described three, 0.5% of respondents described four. Number of respondents for each category of experience is shown in Figure 1. (Double counting was not avoided for respondents who gave more than one experience.) Frequently reported experiences were 'neighbor', 'town', 'TV program' and 'childhood.'

Relationships between experience categories and the condensed attitude scores are also shown in Figure 1. As can be seen in the left panel, the greatest attitude score was seen in the item of 'family.' The next greatest scores were for the categories 'friend', 'volunteer', 'job', and 'visit.' The 'relatives' category was associated with a relatively low condensed score. Minus condensed scores were seen in 'neighbor' and 'town' categories, as well as in the 'error' and 'blank' ones.

Formation of an image of persons with intellectual disability may occur through: 1) 'active connection' or 2) 'passive connection.' A typical 'active connection' would be having a family member with intellectual disability. A typical 'passive association' would be coming across an unacquainted person with intellec-

tual disability in a town. 'Active connection' should result in a positive change in the image of persons with intellectual disability (Wishart & Johnston, 1990). Actually, respondents who stated they had a family member with intellectual disability ('family') scored highly (Figure 1). Experience with a friend's child with intellectual disability, volunteer work or job contact with people with intellectual disability, and visit to a facility for people with intellectual disability can all be taken as a kind of 'active connection.' Respondents who reported such connection also recorded relatively high condensed scores.

On the other hand, 'passive connection' should be connected with lower attitude scores. In fact, 'passive connection' such as coming across in town ('town') or having a neighbor who has a child with intellectual disability ('neighbor'), having seen a relevant TV program ('TV program') was not associated with as high a condensed score as was 'active connection.' By taking into account the finding that formation of images of people with intellectual disability was mainly based on 'passive connection', some part of the reason why attitude amelioration is difficult may be explained.

Self-assessed attitudes and its reason. Respondents who judged themselves as having a more tolerant opinion than the average person (question IIIA) gave mainly seven reasons for it (question IIIB) (see Figure 2). Among them, categories which were associated positively with the condensed attitude score were as follows: 1) the respondent (or a family member) has a disability other than intellectual disability ('another disability'), 2) the respondent knows (or knew) a person who has a family member with intellectual disability (or a person with intellectual disability) and is aware of their situation ('know personally'), 3) the respondent has a person with intellectual disability in the family or among relatives ('family member'), 4) the respondent believes everyone has chance of having a family member with intellectual disability ('equal possibility'), and 5) as a baby the respondent's child had a serious developmental problem with a unfulfilled prognosis of later disability ('seeming retardation'). The respondent imagined him/herself as a person with intellectual disability (or a person who had a family member with intellectual disability) ('picturing oneself') and the respondent during pregnancy had an uneasy feeling without any clear basis that she might have a child with intellectual disability ('anxious feeling') had not so strong relationship with positive attitude of the condensed score.

The reasons listed in 1, 2, 3, and 5 are involuntary/voluntary experiences connected

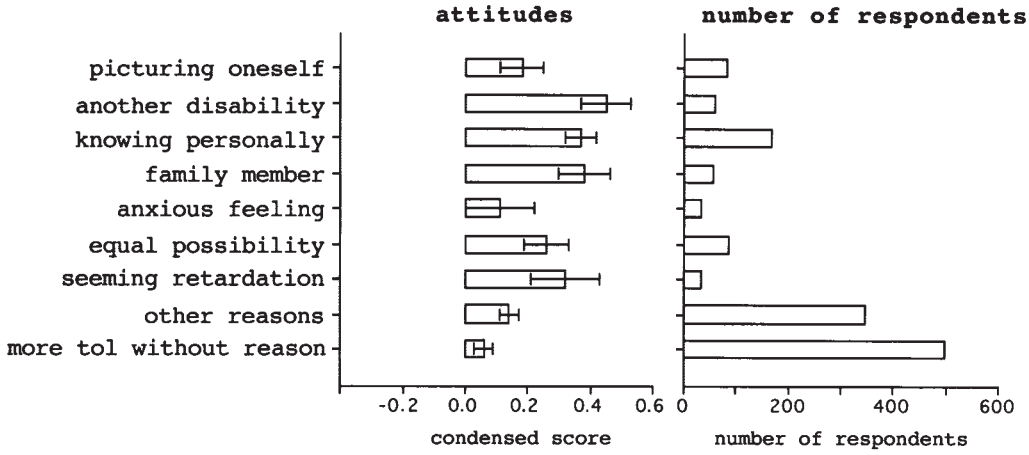

Figure 2. Relation between reasons for having a more tolerant opinion by self-judgment towards persons with intellectual disability and the condensed attitude scores expressed in terms of mean and SE. The numbers of respondents are given in a separate histogram. (See the text for explanation of abbreviated terms. 'more tol without reason': a more tolerant opinion without any stated reason.)

with persons with (intellectual) disability. On the other hand, the idea of 'equal possibility' is information about intellectual disability that is not necessarily connected with actual experiences. One possibly practical way of ameliorating attitude is to provide contact experience. However, we cannot expect that everyone will willingly expose themselves to such contact experience. On the other hand, there is no difficulty in disseminating information that anyone has a chance of having a family member with intellectual disability. Thus, the 'equal possibility' result is an important and hopeful finding for improvement of people's attitude.

Negative experience and self-assessed attitudes. Respondents who judged themselves as less tolerant than the average person gave several reasons for it (see Figure 3). Among them, 'negative experience', i.e., an unpleasant experience with a person with intellectual disability was a prominent group of reasons. Examples of such negative experiences are: "On a bus I was suddenly hit for no obvious reason by a person with intellectual disability." Or, "A person with intellectual disability came too near, uttered a shout, and I had a feeling of fear." Respondents who described themselves as "indeterminate" or "about average" ('same level') and who did not respond to the more/less tolerant question ('blank') were also shown in Figure 3. The left panel of Figure 3 depicts the relationship between condensed scores and having a less tolerant opinion (and 'same level', 'blank' respondents). As can be seen in the panel, respondents who reported a 'negative experience' had a very low condensed attitude score. Respondents who gave other reasons or no reason for having an less tolerant opinion recorded smaller scores than respondents who gave reasons of 'negative experience.'

For more detailed analysis, the 'negative experience' respondents were divided into two groups on the basis of the time of their experience: 1) 'childhood' or 'adulthood'; 2) A feeling of fear without an actual 'negative experience' ('fear feeling') or an actual 'negative experience' ('actual exp') with a person with intellectual disability. An example of 'fear feeling' is: "I fear being with a person with intellectual disability in a closed room because I cannot know what he/she is thinking." An example of an 'actual negative experience' is: "I was touched on my hip by a person with intellectual disability." Some respondents mentioned that persons with intellectual disability might do harm or commit a crime ('harm or crime'). Numbers of respondents for each group are shown in Figure 4. In the

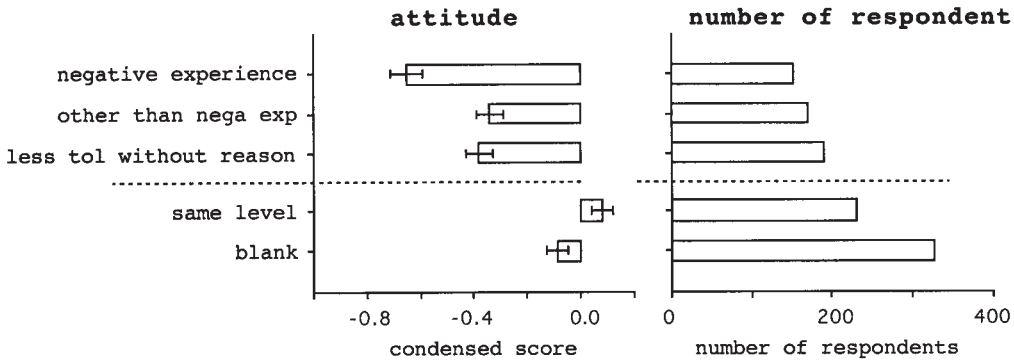

Figure 3. Relation between reasons for having a less tolerant opinion by self-judgment towards persons with intellectual disability and the condensed attitude scores expressed in terms of mean and SE. The numbers of respondents are given in a separate histogram. A smaller condensed score indicates a negative positive attitude toward persons with intellectual disability.
negative experience: an unpleasant experience with a person with intellectual disability.
other than nega exp: reasons other than 'negative experience'
less tol without reason: an less tolerant opinion without any stated reason.
same level: respondents who judged themselves as having an average opinion.
blank: no response on the more/less tolerant question.

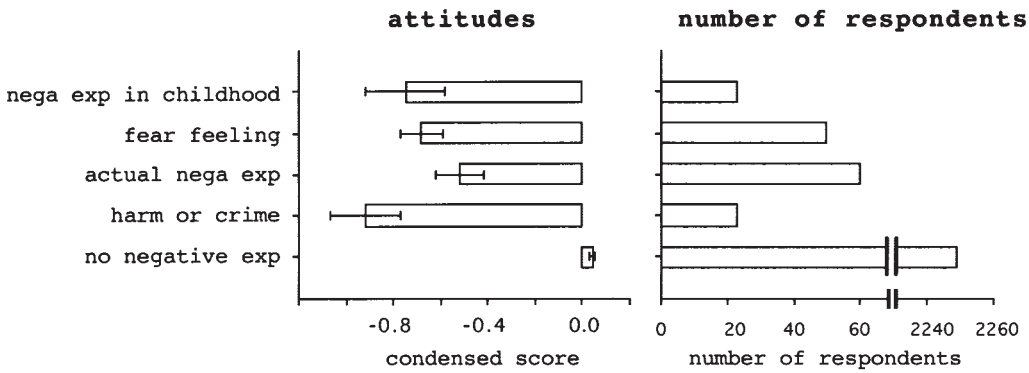

Figure 4. Relation between negative experiences and the condensed attitude scores expressed in terms of mean and SE. The numbers of respondents are given in a separate histogram.
nega exp in childhood: negative experience in childhood (including 'fear feeling' and 'actual exp' as a child).
fear feeling: a feeling of fear without actually negative experience as an adult.
actual nega exp: an actual negative experience as an adult.
harm or crime: mention that some persons with intellectual disability may do harm or commit a crime.
no negative exp: no description of negative experience.

figure, the two types of experience in 'childhood' were pooled due to the small number of respondents ($n = 3$) in 'fear feeling', and are shown in the 'childhood' category. For comparison, results of respondents who did not report having a negative experience were also shown.

As can be seen in the figure, 'negative experience' in childhood (included 'actual exp' and 'fear feeling' but mainly made up with 'actual exp') was associated strongly with the condensed attitude score. In the adult category, the 'fear feeling' group has a lower score than the 'actual negative experience' group. The association of 'negative experience' with lower condensed scores is clear. However, it is also true that not every respondent who had an 'actual negative experience' had small condensed scores. Some respondents (9%) with 'negative experience' had greater condensed scores than 0.5, i.e., a considerable degree of positive attitude. (Scores below 0.5 were recorded by 72% of respondents who did not report 'negative experience.') The reason why some respondents (9%) were not affected so seriously despite their 'actual negative experience' is unclear from the present data. Considering the difficulty of avoiding 'negative experience' in everyday life, this is an important subject for future investigation. It is interesting that even a 'negative experience' in childhood or a 'fear feeling' without such experience could be associated with very low condensed scores. In previous studies, there were no attempts to examine how 'negative experience' with a person with intellectual disability might affect attitudes. However, findings in the present study reinforce the need for studies on the effects of 'negative experience.' The 'harm or crime' respondents had very low condensed scores, but the number of respondents in this group is small ($n = 25$).

Childhood experience, inclusive education for respondent's child, and attitudes. The association (in question I) of 'childhood' experience which formed a perception of persons with intellectual disability was not strong with positive condensed scores (Figure 1). There were frequent references to experiences including interpersonal contact with a child with intellectual disability in response to questions other than I. References to 'childhood' in responses to questions I, IIB, and IIIB were pooled and divided into two types: 1) simple reference to a child with intellectual disability without any concrete description of an interpersonal contact ('simple ref childhood'); for example, one respondent reported, "I saw from a distance pupils with intellectual disabil-

ity in the same elementary school. Since there was no actual contact with them, I got an impression of them from their observed behavior"; and 2) experience of interpersonal contact (mostly through play with children with intellectual disability ('play with'); for example, one respondent noted, "I played with a pupil with intellectual disability many times in my childhood." These items and the relationship of the responses with condensed scores are shown in Figure 5. As seen in the figure, 'play with' respondents had greater condensed scores than had 'simple ref childhood' respondents. This might fit in with some Japanese studies that were skeptical about the efficacy of simple contact experience (see the review by Tachibana & Watanabe, 2002). The importance of early interpersonal contact such as playing with child with intellectual disability is clear from the present results.

In answering questions I, IIB, and IIIB, a few respondents mentioned that their children received inclusive education in a day-nursery or kindergarten ('ref inclusive educa'). It is believed that experience in an inclusive education has a positive effect on the attitude of children without disability. Several relevant studies have been reported (Hastings & Graham, 1995; Townsend, Wilton, & Vakilirad, 1993). In the present investigation, a positive effect was found not on the children but on the mothers of children attending an inclusive education system. Not a few respondents made such comments as; "I had frequently attended events for mothers in the day-nursery. This gave me a chance of making friends with a mother of a child with intellectual disability. I changed my ideas about persons with intellectual disability through having frequent chats with that mother." It is interesting that this effect on attitude was seen especially in day-nursery or kindergarten settings, but not in elementary settings. Perhaps the interest and thus the conversation of mothers whose children are attending to a elementary school might be focused on learning progress of their children, resulting in less frequent mention of pupils with intellectual disability.

Less tolerant respondents' belief and their attitudes. Seventy-six percent of respondents thought that their attitude was more tolerant than (or at least as tolerant as) average. This is a very interesting finding. For more detailed analysis, respondents were divided into quartiles on the basis of condensed attitude score. Then, respondents for each quartile were sorted in terms of response to the more/less tolerant question. Surprisingly, even of respondents in the bottom quartile (i.e., most negative attitude group) on the condensed attitude scores, 43% thought their attitude was more tolerant than (or at least as tolerant as) average. If we drop the 'blank' respondents, the percentage increases to 52%. This finding that so many actually intolerant respondents believe mistakenly they are tolerant might give one reason why the improvement of negative attitude is difficult. Such belief might not give

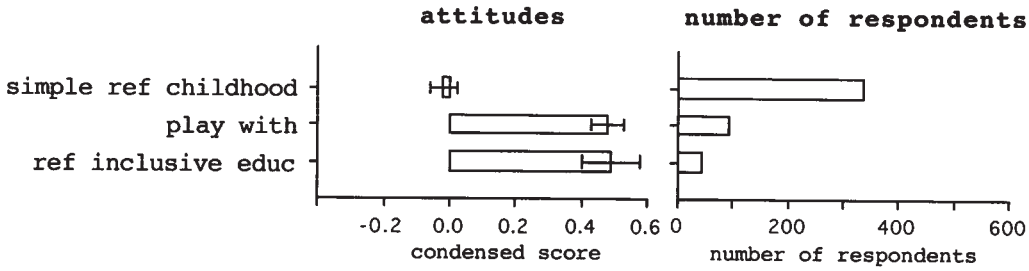

Figure 5. Relation of experience in childhood and inclusive education to the condensed attitude scores expressed in terms of mean and SE. The numbers of respondents for each individual experience was given in a separate histogram.
simple ref childhood: referring simply to their childhood.
play with: interpersonal contact in childhood involving play with a child with intellectual disability.
ref inclusive educ: reference to an inclusive education in which respondents' children attend a day-nursery or kindergarten along with children with disability.

a chance for changing the intolerant opinion, because the belief of their having a tolerant opinion may be part of a desirable ethical standard or norm and, thus, may maintain the opinion.

A limitation of exploratory approach. To find effective ways of improving attitudes toward people with intellectual disability, an 'exploratory' approach was taken. This approach has some methodological weakness. For example, a clear association of a 'play with' experience in childhood with the condensed attitude scores was found (Figure 5). However, the magnitude of the association seen in that figure would probably shrink if it were to be examined by a 'confirmatory' method. Not all respondents who had played with children with intellectual disability will necessarily have described the experience in response to the open-ended question. Probably some respondents who got a strong impression from the experience reported it. To examine more precisely the size of the correlates, a confirmatory question is needed such as, "Did you have an experience of playing with a child with intellectual disability when you were young?"

Despite the weakness of the methodology, however, responses to open-ended questions throw interesting light on formation of attitudes toward people with intellectual disability, though they need to be tested by future confirmatory-type studies.

References

Antonak, R. F., Mulick, J. A., Kobe, F. H., & Fielder, C. R. (1995). Influence of mental retardation severity and respondent characteristics on self-reported attitudes toward mental retardation and eugenics. *Journal of Intellectual Disability Research, 39,* 316–325.

Caruso, D. R., & Hodapp, R. M. (1988). Perception of mental retardation and mental illness. *American Journal on Mental Retardation, 93,* 118–124.

Gottlieb, J., & Siperstein, G. N. (1976). Attitudes toward the mentally retarded: Effect of attitude referent specificity. *American Journal of Mental Deficiency, 80,* 376–381.

Hastings, R. P., & Graham, S. (1995). Adolescents' perceptions of young people with severe learning difficulties: The effects of integration schemes and frequency of contact. *Educational Psychology, 15,* 149–159.

Tachibana, T. (1988). On significance test in non-random assignment of subjects studies: A call for more discussion. *Psychological Reports, 62,* 415–418.

Tachibana, T., & Watanabe, K. (2002). Japanese studies on attitudes toward persons with mental retardation. *Mental Retardation, 40,* 245–251.

Tachibana, T., & Watanabe, K. (2003). Schemata and attitude toward persons with intellectual disability in Japan. *Psychological Reports, 93,* 1161–1172.

Tachibana, T., & Watanabe, K. (2004a). Attitudes of Japanese adults toward persons with intellectual disability: Relationship between attitudes and demographic variables. *Education and Training in Developmental Disabilities, 39,* 109–126.

Tachibana, T., & Watanabe, K. (2004b). Attitudes of Japanese adults toward persons with intellectual disability: Comparisons over time and across countries. *Education and Training in Developmental Disabilities, 39,* 227–239.

Tachibana, T. (in press). Attitudes of Japanese adults toward persons with intellectual disability: Effect of perceptions concerning intellectual disability. *Education and Training in Developmental Disabilities.*

Townsend, M. A. R., Wilton, K. M., & Vakilirad, T. (1993). Children's attitudes toward peers with intellectual disability. *Journal of Intellectual Disability Research, 37,* 405–411.

Williams, R. F. (1986). Perceptions of mentally retarded persons. *Education and Training of the Mentally Retarded, 21,* 13–19.

Wishart, J. G., & Johnston, F. H. (1990). The effects of experience on attribution of a stereotyped personality to children with Down's syndrome. *Journal of Mental Deficiency Research, 34,* 409–420.

Received: 2 June 2004
Revision Received: 28 July 2004
Final Acceptance: 20 January 2005

Education and Training in Developmental Disabilities, 2005, 40(4), 360–367

Teachers' View of Transition Services: Results from a Statewide Survey in South Carolina

Dalun Zhang
Texas A&M University

Joy Ivester
University of South Carolina

Antonis Katsiyannis
Clemson University

Abstract: This study examined high school and middle school transition practices from the perspective of special education teacher leaders. Results indicated that the majority of schools offer functional curricula, social skills training, and an array of school-based functional learning opportunities. However, a significant number of schools did not provide adequate employment skills development and functional assessment. In addition, agency involvement in transition planning was inadequate.

Transition from school to work was identified as a major federal priority in special education since the early 1980s. The Office of Special Education and Rehabilitation Services (OSERS) through transition systems change initiatives, such as The School-to-Work Initiative (Guy & Schriner, 1997), has served as the impetus for systemic changes in transition practices for post school settings, including work, postsecondary education and community living (Johnson & Halloran, 1997; Katsiyannis, deFur, & Conderman, 1998). States which have participated in systems change initiatives reported an increase (a) in awareness of transition needs and issues, (b) in participation of students with disabilities and parents in transition activities, (c) in collaborative relationships between school and community agencies, and (d) in policy development to support better transition services and outcomes (Guy & Schriner, 1997). Further, federal funding for these initiatives was found to be the best predictor for improvement in transition practices though "transition compliance efforts" appear to be focused more on the process (i.e., paperwork) than the provision of transition services (Baer, Simmons, & Flexer, 1996). Nonetheless, special education teachers, school administrators, counselors, and agency personnel indicated that transition topics were discussed more frequently and significant progress had been made in transition planning and interagency collaboration (Roessler, 1996).

Schools generally report compliance with IDEA transition requirements with the exception of the requirement to reconvene the Individualized Education Program (IEP) team if transition services could not be provided as planned. School personnel must ensure that students with disabilities fully access and benefit from the general education curriculum and leave our school systems prepared to successfully participate in postsecondary education, enter meaningful employment, live independently in communities, and pursue lifelong learning opportunities (Johnson, Stodden, Emanuel, Luecking, & Mack, 2002). Early transition planning, student/parent involvement and ownership of plans, age and goal appropriate environments, and a current directory of transition resources have been emphasized as essential elements in effective practices (Afee & Greenawalt, 2001; see also McDonnell, Hardman, & McDonnell, 2003). Additional components include a focus on community outcomes, and interagency collaboration (Morningstar, Kleinhammer, & Lattin, 1999). Although school districts always retain responsibility for school-aged transition services (McAfee & Greenawalt), the existence

Correspondence concerning this article should be addressed to Dalun Zhang, Department of Educational Psychology, 4225 Texas A&M University, College Station, TX 77843-4225.

of a school-based interagency transition team was the strongest predictor of compliance and best practice (Mahan & Baer, 2001).

A variety of models have been developed and implemented to facilitate successful transition. For example, the Transition Service Integration Model combines the resources of public schools, rehabilitation, and developmental disability systems during the student's last year at school (Certo et al., 2003); Project RENEW provides for comprehensive case coordination for the participants' ongoing education, employment, social/emotional development, and community adjustment (participants made improvements in high school completion, enrollment in postsecondary education programs, hours worked per week, and in their hourly wages) (Malloy, Cheney, & Cormier, 1998); Teaching All Students Skills for Employment and Life involves a multilevel interagency transition planning process which includes three levels - a community-level team, school-level team, and individual-level team (evaluation data indicated consumer satisfaction) (Aspel, Bettis, Quinn, Test, & Wood, 1999).

In addition, specific practices, such as an emphasis on vocational training and interagency collaboration, have resulted in significant outcomes for individuals with disabilities. Harvey (2002) found that individuals with disabilities who participated in vocational education while in high school earned more wages than their peers with disabilities who did not participate in vocational education. Karpinski, Neubert, and Graham (1992) found that graduates had worked proportionately more time since high school than dropouts and had been employed in their current job more than twice as long (phone interviews were conducted 21 and 28 months after leaving school). Wehman (2002) also emphasized the need for students to secure competitive employment before leaving school by establishing partnerships between schools and public agencies and private sector as well as expanded postsecondary education opportunities. Financial incentives for public and private colleges that enroll, support, and graduate students with disabilities are perhaps a means to address this issue (Wehman).

A persistent challenge in planning and implementing transition services has been the capacity and training of special education teachers and other school based professionals (see Johnson & Rusch, 1993). Bull, Montgomery, and Beard (1994) pointed out that teacher education programs in colleges and universities must be charged with the responsibilities of adequately preparing special education teachers to plan, implement, and evaluate transition programs. Knott and Asselin (1999) investigated perceptions of 214 secondary teachers of students with mild disabilities in terms of (a) their knowledge of transition, (b) family and student involvement in transition, and (c) importance of transition. Findings indicated that teachers understood the problems, issues, concepts, and definitions related to transition; and placed high importance on family and student involvement in transition planning. However, most teachers had little knowledge of adult agencies' eligibility requirements or roles of these agencies in transition planning; demonstrated low knowledge and involvement in linkages with adult services; and did not know how to accomplish linkages between secondary school activities and adult options. Based on these findings, Knott and Asselin suggest that preservice and inservice programs need to address the lack of knowledge and involvement in interagency collaboration.

The current investigation examined high school and middle school transition practices from the perspective of special education teachers. Surveys were utilized to determine the types and levels of transition services available for students with special education needs in the State of South Carolina.

Method

Participants

The population sampled was the department leader or special education teacher of each middle school and each high school in South Carolina. Participants included 105 middle and high school teachers across South Carolina who completed and returned the survey. Among the participants, 95 (91%) were special education teachers, 2 (2%) were vocational teachers, and 8 (7%) identified themselves as other, including special education

consultant, transition facilitator, employment specialist, counselor, speech language pathologist, placement chair, and special needs coordinator. The types of settings that they taught in included resource (n = 37, 35%), self-contained (n = 43, 41%), inclusion (n = 1, 1%), itinerary (n = 3, 3%), multiple settings (n = 13, 12%), and other (n = 13, 12%).

Instrument

The second author and staff of Project SIGHT (South Carolina Transition Systems Change Project) developed the survey using the format of an earlier survey conducted by Project SIGHT in 1997. The earlier survey was developed with input and reviews from special education coordinators and personnel representing the University of South Carolina Center for Excellence, the SC Department of Education, local education agencies, the Department of Vocational Rehabilitation and the Developmental Disabilities Council, Office of the Governor. Survey items were generated based on a review of the literature on valued transition practices, which addressed service coordination, key stakeholders' involvements, linkages to service agencies, adult service agencies' participations, work-based experiences, school-based experiences, transportation, and transition service categories. Items in the survey were divided into three sections. Section I collected data on participants' background information; which included their professional role, instructional setting, and the grade levels and number of students they had in their classes. Section II asked respondents to report how services were provided in their district. Questions included "who coordinates transition programs in the district, transition coordinator/specialist's responsibilities, regular participants of IEP/transition plan meetings, and agencies that provided information or assistance to them. In Section III, respondents were asked to check services and experiences that students received (see Table 1). These included work-based experiences, school-based skills training, types of businesses that students were employed, and percentages of time students were employed. In addition, respondents rated how well their district addressed each of 15 transition service categories. These categories are listed in Table 3.

TABLE 1

Types of Work-Based Experiences or Information that Students Received Assistance

Type of Experiences or Information	*Number*	*Percent*
School-based work experiences	86	81.9
Receiving career information	74	70.5
Vocational/occupational courses	73	69.5
Job shadowing	70	66.7
Assistance form job coach	59	56.2
Volunteer work	53	50.5
Service learning	50	47.6
Community-based training	50	47.6
Supported employment	38	36.2
Internship	27	25.7
Other	6	5.7

Data Collection and Analysis

The staff of Project SIGHT mailed surveys to each of the 206 high schools, 227 middle school and four junior high school special education department coordinators or lead teachers in the state of South Carolina. Department leaders/teachers were asked to complete the survey and return it within a one-month period in a self-addressed stamped envelope included. At the conclusion of the one-month window, follow-up letters were mailed to each of the survey recipients to encourage follow-through. Returned surveys were coded and entered into SPSS (Statistical Package for Social Sciences).

Three types of statistical analyses were conducted. First, frequency and descriptive statistics were calculated to obtain an overall summary of the data. Second, crosstabulation analyses were conducted to examine relationships between teachers' responses and their instructional settings and between teachers' responses and grade levels. Third, inferential analyses were conducted to examine teachers' ratings on their school's transition services in relation to their instructional settings and grade levels.

Results

How Were Transition Services Provided in the District?

Respondents were asked to identify the responsible person who coordinated transition programming in their district/school. Responses varied, including 46 (43.8%) transition coordinator, 23 (21.9%) multiple personnel, 18 (17.1%) special education director/coordinator, 13 (12.4%) special education classroom teacher, 2 (1.9%) guidance counselor, and 3 (2.9%) other. According to these respondents, their district's transition coordinator carried multiple responsibilities. These included establishing employment site for students with disabilities (n = 50, 47.6%), facilitating student job placement (n = 49, 46.7%), facilitating student assessment (n = 41, 39%), providing supported employment services (n = 41, 39%), facilitating teacher training for special education transition curriculum and services (n = 36, 34.3%), and others (n = 8, 7.6%).

Special education teachers were checked by 87.6% (n = 92) of the respondents as usually present at students' IEP/transition planning meetings, followed by parents (n = 90, 85.7%), general education teachers (n = 84, 80%), the student (n = 83, 79%), guidance counselors (n = 68, 64.8%), and job coaches (n = 33, 31.4%). In terms of agencies' involvement, more than half (n = 72, 68.6%) of the respondents checked Vocational Rehabilitation Department, 58 (55.2%) Department of Disabilities and Special Needs, 33 (31.4%) Department of Mental Health, and 29 (27.6%) checked Department of Social Services. Other agencies that were reported included, Continuum of Care (n = 18, 17.1%), South Carolina Services Information System (n = 11, 10.5%), Social Security Administration (n = 10, 9.5%), and Health Department (n = 9, 8.6%).

Transition Services and Experiences Students Received

Respondents were asked to check the types of work-based experiences or information that their students with special needs received assistance with through the school district. These data are summarized in Table 1. As shown in Table 1, school-based work experiences was the most common experiences students with special needs received from their school district, followed by receiving career information, and vocational/occupational courses. Internship and supported employment were the two least common experiences that students received from their school district. Respondents were also asked to check the types of school-based training skills their students received at school. These data are summarized and ranked from the most common to the least common in Table 2. As shown in Table 2, functional skills curriculum was overwhelmingly identified followed by social skills curriculum and independent living skills curriculum.

Regarding what types of businesses were the students employed, more than half (n = 69, 65.7%) checked grocery store, followed by fast food, (n = 65, 61.9%), restaurant (n = 58, 55.2%), service industry (n = 38, 36.2%), construction (n = 30, 28.6%), and manufacturing (n = 24, 22.9%). Farming was the least reported business (n = 9, 8.6%). Sixty-four respondents answered the questions "How does your district provide transportation for students to and from work/training?" Of those who responded, 5 (7.8%) reported using teacher personal vehicle, 22 (34.4%) reported using activity bus, 9 (14.1%) reported using

TABLE 2

Types of School-Based Training Skills Students Received at School

School-Based Training Skills	*Number*	*Percent*
Functional curriculum (daily living skills, job seeking/keeping skills, etc.)	94	89.5
Social skills curriculum	74	70.5
Independent living skills curriculum	72	68.6
Vocational/occupational skills curriculum	71	67.6
Assessment (interviews, career occupational profile, etc.)	69	65.7
School-to-work education	63	60.0
Employability skills curriculum	60	57.1
Other	2	1.9

school bus, and 28 (43.8%) reported using other types of transportations.

Quality of School/District Services

Respondents were asked to rate how well their school or district addressed each of 15 statements regarding transition education/services. The rating scale consisted of five choices ranging from 0 to 4: 0 = not addressed by school, 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree. The 15 statements and descriptive statistics for each statement are summarized in Table 3. As shown in Table 3, individualized education/transition plan, vocational/occupational skills curriculum, and functional curriculum seem to be addressed better than other issues. Six transition issues were reportedly not addressed by over 10 percent of the districts/schools. These included internship, job coaching, supported employment, community-based instruction service learning experiences, and employability skills.

To examine whether respondents from different school levels rated differently, we conducted a one-way analysis of variance (ANOVA). The dependent variable was rating total (sum of all 15 ratings). The independent variable was school level (including 21 middle school, 54 high school, and 10 multiple level). The one-way ANOVA yielded an F value of 7.50 ($p < .005$), which indicated a significant difference. Further multiple comparison analysis using Scheffe procedure detected differences existed between high school and multiple level (mean difference = 14.11, $p < .005$) and between middle school and multiple level (mean difference = 11.84, $p < .01$). In addition, we conducted a one-way ANOVA to examine whether respondents' ratings were different based on their instructional settings (self-contained, resource room, and others). Results of the analysis did not reveal any significant differences.

Discussion

Research demonstrates that existence of specific transition practices appear to be directly related to student success in postsecondary environments (Harvey, 2002; Karpinski et al.,

TABLE 3

Respondent's Ratings on How Well Their School/District Addressed 15 Transition Issues

Transition Issue	*Number and Percentage on Each Response*				
	SA	*A*	*D*	*SD*	*NO*
Individualized education/transition plan (n = 98)	29 (30)	55 (56)	8 (8)	2 (2)	4 (4)
School-to-work education (n = 98)	18 (18)	51 (52)	16 (16)	6 (6)	7 (7)
Functional curriculum (n = 99)	24 (24)	52 (53)	17 (17)	2 (2)	4 (4)
Support employment (n = 96)	13 (14)	38 (40)	17 (18)	9 (8)	19 (20)
Community-based instruction (n = 97)	18 (19)	44 (45)	18 (19)	4 (4)	13 (13)
Assessment (n = 97)	21 (22)	55 (57)	13 (13)	6 (6)	2 (2)
Social skills curriculum (n = 98)	21 (21)	48 (49)	23 (24)	1 (1)	5 (5)
Independent living skills curriculum (n = 97)	20 (21)	53 (55)	17 (17)	1 (1)	6 (6)
Vocational/occupational skills curriculum (n = 97)	25 (26)	46 (47)	12 (12)	8 (8)	5 (5)
Employability skills curriculum (n = 97)	20 (21)	40 (41)	23 (24)	4 (4)	10 (10)
Internship (n = 94)	9 (10)	32 (34)	25 (27)	6 (6)	22 (23)
Job coaching (n = 98)	23 (23)	30 (31)	19 (19)	6 (6)	20 (21)
Service learning experiences (n = 97)	17 (18)	44 (45)	22 (23)	3 (3)	11 (11)
Providing career information (n = 96)	22 (23)	49 (51)	11 (12)	9 (9)	5 (5)
Transition education for compliance to special education legislation (n = 94)	17 (18)	52 (55)	15 (16)	3 (3)	7 (8)

Note. SA = strongly agree, A = agree, D = disagree, SD = strongly disagree, NO = not addressed. Numbers in parentheses are percentages based on number of responses.

1992). The purpose of this investigation was to identify whether, from the perspectives of special education department leaders/teachers, schools in the State of South Carolina are applying transition best practices as generally defined by researchers and practitioners (e.g., Guy & Schriner, 1997; Afee & Greenawalt, 2001; Mahan & Baer, 2001). The outcomes of the study provide a snapshot of transition practices in South Carolina.

Participant Background Information

Results indicated that most school districts have identified someone to serve as a transition point-person. Identifying such a person is significant because it is an indicator of the relative importance school districts placed on transition and the visibility of such a position. These individuals often assume multiple roles. For instance, in addition to serving as the technical assistance and training staff for the district, they are also often responsible for direct student services such as job coaching and development and student assessment. The assumption of multiple roles may be especially beneficial when responsibilities cross special education and general education lines. For example, a transition coordinator who also serves as a district vocational coordinator often has greater access to general education programs and resources.

District Service Provision

Findings from this survey also point to several areas of concern. Specifically, level of student participation in transition planning appears inadequate. Survey respondents indicated that only 79% of students attended such meetings. Field, Martin, Miller, Ward, and Wehmeyer (1998) concluded that a student's active participation in the IEP meeting results in high degree of engagement. Wehmeyer (1998) points out that (a) active participation in the IEP process makes the student more motivated to pursue goals that they have helped select and (b) students who are involved in setting goals have more positive outcomes related to achieving those goals than with goals selected by others. Additionally, it seems that agency representatives are attending only a select number of transition planning meetings. Although research constantly emphasizes the importance of rehabilitation counselors' involvement in transition planning (e.g., Agran, Cain, & Cavin, 2002), Vocational Rehabilitation is present at meetings in barely more than half of the schools. More disturbingly, the Department of Disabilities and Special Needs, Mental Health, and other agencies are absent from the transition process in the majority of the schools. This is certainly an area that needs to be addressed in order to bridge the gap between high school and postsecondary environments. Research findings have consistently documented that effective transition programs involve a strong collaboration component whereas lack of interagency collaboration often becomes a barrier to effective transition planning and services (Kohler, 1998). Interestingly enough, though Vocational Rehabilitation presence was noted in more than half of the schools, students do not appear to be receiving supported employment services (community-based) during their school years. However, participants noted that students are being provided with school-based work experiences in an overwhelming majority (81.9%) of schools.

In sum, the majority of schools in South Carolina report that they are not providing adequate employment skill development for their students. This finding is most serious when research indicates time and time again that while employment levels for adults with disabilities fall well below those of other citizens (Presidential Task Force on Employment of Adults with Disabilities, 1999) work experience and placement during high school is directly correlated to post-school employment success (Benz, Yovanoff, & Doren, 1997; Rabren, Dunn, & Chambers, 2002). Further, more than half of schools have students involved in the food industry (i.e., restaurant or grocery stores). This limited array of employment options may be the result of students being placed in available, convenient jobs in lieu of matching individual student interests to specially selected positions.

In contrast, schools in South Carolina are successful in providing school-based functional skills development opportunities to students. Nearly 90% of schools reported that they offer functional curricula for students and over 70% report offering a social skills

curriculum. In addition, over two thirds report offering an array of school-based functional learning opportunities (i.e., independent living skills, vocational and vocational related skills, and daily living skills). Unfortunately, about one-third failed to indicate that they provide functional assessments, a critical step in providing an individualized, appropriate transition program for students with disabilities.

Findings from this study should be viewed with caution as they represent transition practices in South Carolina. Findings also represent the perspective of primarily special education teachers and only a fraction (about 25%) of teachers participated in the study. In addition, because only lead teachers were targeted, it is possible that an underlying level of responsibility for transition programming could have skewed responses. Further research is needed to ensure that alternate perspectives from other professionals involved in the process (transition coordinators, school-based and agency personnel), parents, and students are investigated so that an accurate and comprehensive view of transition practices in South Carolina is possible.

In summary, findings regarding transition practices in South Carolina (despite the study's limitations) allow for preliminary observations. Results indicated that the majority of schools offer functional curricula, social skills training, and an array of school-based functional learning opportunities. However, a significant number of schools did not provide adequate employment skills development and functional assessment. In addition, agency involvement in transition planning was inadequate. Finally, broadening the base of survey participants to additional stakeholders involved in the transition process along with improved levels of participation will allow for a more comprehensive (and potentially more accurate) view of transition services.

References

Agran, M., Cain, H. M., & Cavin, M. D. (2002). Enhancing the involvement of rehabilitation counselors in the transition process. *Career Development for Exceptional Individuals, 25,* 141–155.

Aspel, N., Bettis, G., Quinn, P., Test, D., & Wood, W. (1999). A collaborative process for planning transition services for all students with disabilities. *Career Development for Exceptional Individuals, 22,* 21–42.

Baer, R., Simmons, T., & Flexer, R. (1996). Transition practice and policy compliance in Ohio: A survey of secondary special educators. *Career Development for Exceptional Individuals, 19,* 61–71.

Benz, M. R., Yovanoff, P., & Doren, B. (1997). School-to-work components that predict post-school success for students with and without disabilities. *Exceptional Children, 63,* 151–164.

Bull, K. S., Montgomery, D., & Beard, J. (1994). Teacher competencies for transition programs as reported by state directors of special education. *Rural Special Education Quarterly, 13*(4), 10–16.

Certo, N. J., Mautz, D., Pumpian, I., Sax, C., Smalley, K., Wade, H., et al. (2003). A review and discussion of a model for seamless transition to adulthood. *Education and Training in Mental Retardation and Developmental Disabilities, 38,* 3–17.

Field, S., Martin, J., Miller, R., Ward, M., & Wehmeyer, M. (1998). *A practical guide for teaching self-determination.* Reston, VA: The Council for Exceptional Children.

Guy, B., & Schriner, K. (1997). Systems in transition: Are we there yet? *Career Development for Exceptional Individuals, 20,* 141–164.

Harvey, M. W. (2002). Comparison of postsecondary transitional outcomes between students with and without disabilities by secondary vocational education participation: Findings from the national education longitudinal study. *Career Development for Exceptional Individuals, 25,* 99–121.

Johnson, D., & Halloran, W. (1997). The federal legislative context and goals of the state systems change initiative on transition for youth with disabilities. *Career Development for Exceptional Individuals, 20,* 109–121.

Johnson, J., & Rusch, F. (1993). Secondary special education and transition services: Identification and recommendations for future research and demonstration. *Career Development for Exceptional Individuals, 16,* 1–18.

Johnson, D., Stodden, R., Emanuel, E., Luecking, R., & Mack, M. (2002). Current challenges facing secondary education and transition services: What research tells us. *Exceptional Children, 68,* 519–542.

Karpinski, M. J., Neubert, D. A., & Graham, S. (1992). A follow-up study of postsecondary outcomes for graduates and dropouts with mild disabilities in a rural setting. *Journal of Learning Disabilities 25,* 376–385.

Katsiyannis, A., deFur, S., & Conderman, G. (1998). Transition services–systems change for youth with disabilities? A review of state practices. *Journal of Special Education, 32,* 55–61.

Knott, L., & Asselin, S. B. (1999). Transition com-

petencies: Perception of secondary special education teachers. *Teacher Education and Special Education, 22,* 55–65.

Kohler, P. D. (1998). Implementing a transition perspective of education: A comprehensive approach to planning and delivering secondary education and transition services. In F. R. Rusch & J. G. Chadsey (Eds.), *Beyond high school: Transition from school to work* (pp. 179–205). Belmont, CA: Wadsworth Publishing.

Malloy, J., Cheney, D., & Cormier, G. (1998). Interagency collaboration and the transition to adulthood for students with emotional or behavioral disabilities. *Education and Treatment of Children, 21,* 303–320.

McAfee, J., & Greenawalt, C. (2001). IDEA, the courts, and the law of transition (Individuals with Disabilities Education Act). *Preventing School Failure, 45*(3), 102.

McDonnell, J., Hardman, M. L., & McDonnell, A. (2003). *Introduction to people with severe disabilities* (2nd ed.). Boston: Allyn & Bacon.

McMahan, R., & Baer, R., (2001). IDEA transition policy compliance and best practice: Perceptions of transition stakeholders. *Career Development for Exceptional Individuals, 24,* 169–184.

Morningstar, M., Kleinhammer, T., P., J., & Lattin, D. (1999). Using successful models of student-centered transition planning and services for adolescents with disabilities. *Focus on Exceptional Children, 31*(9), 1–19.

Presidential Task Force on Employment of Adults with Disabilities. (1999). Report from the subcommittee on expanding employment opportunities for young people with disabilities. Washington, DC: Author.

Rabren, K., Dunn, C., & Chambers, D. (2002). Predictors of post-high school employment among young adults with disabilities. *Career Development for Exceptional Individuals, 25,* 25–40.

Roessler, R. (1996). An evaluation of transition activities by demonstration site professionals. *Rural Special Education Quarterly, 15*(2), 43–48.

Wehman, P. (2002). A new era: Revitalizing special education for children and their families. *Focus on Autism and Other Developmental Disabilities, 17,* 194–198.

Wehmeyer, M. L. (1998). Student involvement in education planning, decision making, and instruction: An idea whose time has arrived. In M. Wehmeyer & D. Sands (Eds.), *Making it happen: Student involvement in educational planning, decision-making and instruction* (pp. 3–23). Baltimore: Brookes.

Received: 10 March 2004
Initial Acceptance: 23 April 2004
Final Acceptance: 25 October 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 368–376

Using Self Evaluation to Improve Student Teacher Interns' Use of Specific Praise

Cassandra L. Keller
Lynn University

Michael P. Brady
and Ronald L. Taylor
Florida Atlantic University

Abstract: This study investigated effects of an audiotape self evaluation intervention on the instructional behavior of three student teacher interns in classrooms for students with various developmental disabilities. A multiple baseline design was used to evaluate of the intervention. Results indicated a positive effect on all interns' use of specific social praise. Generalization probes indicated two of three interns increased their use of specific social praise in non-targeted content areas. Maintenance probes indicated all interns' average use of specific social praise was well above baseline. Implications for teacher educators and classroom teachers of students with developmental disabilities are discussed.

Self evaluation is a process in which one makes judgments about the adequacy and effectiveness of performance for the purpose of self improvement. It is the most common form of evaluation used by teachers to improve practice (Airasian, Gullickson, Hahn, & Farland, 1995). The function of self evaluation is to help teachers identify and make decisions about the strengths and weaknesses of their practice, with the intent of improving it (Airasian & Gullickson, 1997). Self evaluation has been identified as a method for promoting increased use of effective practices in teachers (Bullard, 1998; Kilbourn, 1991; Stronge, 1997; Sutherland & Wehby, 2001).

Although research has indicated that data-based self evaluation procedures have been extremely effective in helping practicing teachers increase their use of effective teaching behaviors, very few self evaluation studies have been conducted with student teacher interns. Of the small number of self evaluation studies that have been conducted with teacher interns, most have used rating scales or questionnaires to assess or promote reflections or general perceptions of their teaching (Griffin & Kilgore, 1995; Wheeler & Knoop, 1982). Many educators (Buchanan & Jackson, 1998; Freiberg & Waxman, 1990; Greis, 1986; Kagan & Tippens, 1991; Speer & Zoellick, 1974) have discussed the value of using self evaluation techniques, and how interns might benefit from a more data-based approach to evaluating behavior. This includes attention to accuracy in collecting specific data on one's own teaching, as well as using these data to make decisions. Frequently, initial attempts at self evaluation do not correspond with more objective, external evaluations (Dougall & Brady, 1998). For example, Briggs, Richardson, and Sefzik (1986) and Wheeler and Knoop found that interns' self evaluations of their performance tended to be more positive than their supervisors' ratings. This suggests that interns may not be accurate in self evaluating their performance without *specific* data on their own instructional behavior.

Correspondence concerning this article should be addressed to Cassandra L. Keller, Lynn University, Ross College of Education, 3601 N. Military Trail, Boca Raton, FL 33431. E-mail: ckeller@lynn.edu

Specific Praise as an Effective Teaching Behavior

Praise is an instructional behavior that has been shown to be a critical component of effective instruction (Emmer, 1988; Englert, 1984; Hancock, 2000; Sutherland, Wehby, & Yoder, 2002). Yet, studies consistently report that teachers provide an extremely low rate of praise to students in general education and special education classes (Brady & Taylor, 1989; Gable, Hendrickson, Young, Shores & Stowitschek, 1983; Gunter & Coutinho, 1997). Teacher verbal praise has been reported to occur at the rate of less than

five times per hour per student in general education classrooms and less than two times per hour per student in classrooms for students with disabilities (Brophy, 1981; Van Acker, Grant, Van & Henry, 1996; Wehby, Symons, & Shores, 1995). This indicates that students placed in special education classrooms often receive less praise than students in general education classrooms.

Numerous studies indicate teacher praise increases student motivation, task engagement, and actual learning. Sutherland, Wehby, and Copeland (2000) indicated that students' task engagement increased when teacher praise increased, and decreased when praise was decreased. They found that as teachers increased their use of praise, accuracy of student academic responding improved, resulting in higher academic achievement. Hancock (2000) found that students exposed to verbal praise persisted longer at tasks than a control group that received no praise. He also found it to be a practical way of affecting students' motivation to engage in behaviors associated with learning.

In a review of 30 years of empirical research on praise and reprimands, Beaman and Wheldall (2000) determined the typical rates used by teachers in elementary through secondary schools. They found students' academic behavior is praised more often than social behavior and teachers respond more frequently to inappropriate social behavior than to appropriate social behavior. They suggested that teachers frequently fail to take advantage of praise to manage the social behavior of their students.

These studies and others make a strong case for teachers' use of frequent, specific praise to increase students' involvement in lessons, their motivation to learn, and their academic performance. Thus, research is needed on techniques that may promote and maintain increased use of effective teaching behaviors such as specific praise. Given the impact student teaching may have on future behavior (Guyton & McIntyre, 1990), it is clear that studying ways to promote interns' use of effective teaching behavior such as praise is important.

This study examined effects of a self evaluation intervention using audiotape samples of interns' instructional behavior on the frequency of their (a) specific social praise in classrooms for students with disabilities, (b) specific social praise in non-targeted content areas, and (c) specific social praise after the intervention ended.

Method

Participants and Setting

Three female undergraduate students majoring in Exceptional Student Education at a regional university in South Florida participated. All were completing the student teaching portion of their programs in elementary school settings. They were also enrolled in a special program called Accelerated Induction into Teaching (AIT) in which they served full-time as substitute teachers while completing student teaching requirements. Each intern was assigned a master mentor for 2-3 hours per day for support, and a university supervisor who evaluated the intern's performance. All interns taught in special education self contained classrooms with students in grades K-5 with various developmental disabilities.

Interns were European Americans who ranged in age from 24-36 years and had similar prior knowledge and experience with persons with disabilities. Intern 1 (Tracey) had mentored youth with disabilities. Intern 2 (Candice) had worked in a group home for teenagers with disabilities and had a brother with disabilities. Intern 3 (Susan) was a teaching assistant in a classroom with students with disabilities.

Experimental Design

A multiple baseline design across subjects (Baer, Wolfe, & Risley, 1968) was used to evaluate the intervention. The target behavior was frequency of specific social praise. Specific social praise was defined as verbal statements demonstrating approval of a student's social response that specified the behavior being praised. The treatment was the self evaluation procedure which consisted of (a) a one-time scripted training session with each intern, and (b) the intern's participation in daily data-collection and graphing activities.

Procedure

To allow interns and students to adjust to one another, no experimental procedures were implemented during the first two weeks of

student teaching. During the 3rd week, interns met with the researcher to receive the equipment for the study. This included; a micro audiotape recorder, a lapel microphone, a box for tapes, and 10 blank pre-dated audiotapes. Interns were then asked which *teacher-led content area* they wished to target for the study. Teacher-led instruction was defined as instruction to either a small or large group of students that consisted of at least 15 minutes of interaction between intern and students. Teacher-led instruction did not include seatwork or one-to-one instruction. Tracey targeted the content area she called "calendar time." Calendar time was a time when her students orally practiced functional math and oral language skills. Candice targeted reading and Susan targeted math.

Interns were asked to record their instruction during the targeted content area the following week, and to continue to record the same targeted content area each day until the study was over. During the first three days, no data were evaluated by the interns or researcher. This allowed interns and students to acclimate to the audiotape recorder and data collection procedure (turning recorder on and off, etc.). It also reduced the likelihood that data collection would have a reactive effect on either interns or students (Richards, Taylor, Ramasamy, & Richards, 1999).

During baseline, each intern was asked not to listen to the audiotapes. Tapes were collected regularly (daily or every other day) to assure the request. During intervention, interns did have access to their data tapes, consistent with intervention procedures described under Intervention. When the intervention was removed to assess maintenance, interns were asked once again to refrain from listening to their data tapes, and tapes were again collected.

To ensure random data collection, time intervals of 5 min each were written on index cards. Time intervals started at 0-5 min, then 1-6 min and continued up to 10-15 min. Cards were placed into a bag and drawn each day to determine the interval for data collection.

Once the segment of an intern's lesson was determined, the researcher recorded frequency of the target behavior since the time of each lesson (5 min) was consistent throughout the study. During baseline, data were collected for at least five days and until at least one intern showed a stable trend. Baseline was considered stable if it did not vary more than 50% from its mean (Alberto & Troutman, 2003). Once baseline became stable, intervention began with the intern with the lowest mean frequency of specific social praise during baseline. The intern with the second lowest mean began intervention next, and so on.

Intervention

Intervention for each intern started with an individual training session, including a 60-minute scripted lesson that summarized the nine steps for self evaluation adapted from Airasian and Gullickson (1997) and Sutherland (2000):

1. The intern was asked to make a prediction about her frequency of specific social praise in a 5-min period of teacher-led instruction. According to Airasian and Gullickson (1997), self evaluation is most effective when teachers make predictions about their performance before the data are shared with them.
2. The researcher shared with each intern the average frequency of specific social praise statements per 5 min of instruction from baseline audiotapes.
3. The intern was asked to target the frequency of specific social praise statements for change.
4. The intern was provided with characteristics of effective specific praise and the potential benefits of using effective praise in the classroom.
5. The intern was instructed how to use the recorder and protocol for recording frequency of social praise statements.
6. A training audiotape created by the researcher was then coded by both the intern and the researcher. Results were compared and 100% agreement was obtained with each intern on frequency of specific social praise statements.
7. The intern was asked to set a personal goal for frequency of specific social praise per 5 min of teacher-led instruction. The researcher and intern then brainstormed strategies for use during instruction to help meet the goal. The intern was asked to use the strategy or strategies each day to help reach the personal goal.

8. The intern was asked to record the targeted content area during teacher-led instruction and listen to a 5 min sample of her instructional behavior daily. Gunter and Reed (1996) suggested that a five-minute audiotape sample of instruction was sufficient in providing insight into instructional behavior of teachers.
9. A graph was prepared for each intern. Graphs contained the intern's baseline data and a goal statement to serve as a reminder of the interns' target goal. The intern was asked to graph her frequency of specific social praise statements each day until the intervention ended.

A cue card (a visual reminder) was provided to each intern to give her a visual prompt to use effective specific praise frequently. Interns were asked to display the cue cards somewhere in the classroom where they were able to view them while providing instruction.

Generalization and Maintenance

Mentors recorded each intern's instruction during non-targeted content areas to assess generalization of praise to lessons that had not been targeted. Non-targeted content areas were selected by the mentor when the intern provided teacher-led instruction. These audiotapes were analyzed using the same procedures described earlier. Frequency of praise statements indicated whether generalization to non-targeted content areas occurred.

Mentors also recorded each intern's instructional lessons during the *targeted* content area 3½ weeks after the self evaluation intervention ceased. Frequency of each intern's use of specific social praise statements indicated whether interns maintained improvements over time.

Interobserver Agreement

The researcher and a trained graduate student in special education independently recorded frequency of praise statements in 21% of the taped lessons randomly selected across baseline, intervention, generalization and maintenance. The graduate student was naïve to which condition the tape involved. Percent agreement was calculated by dividing the smaller number of praise statements by the larger number and multiplying by 100. Mean agreement across all conditions for Tracey, Candice, and Susan was 94%, 96%, and 93% respectively.

Results

As indicated in Figure 1, all interns' praise statements increased when the intervention was introduced.

Tracey. Frequency of specific social praise increased from baseline (M = 0.3, range 0-1) to intervention (M = 4.8, range 1-13) for Tracey. A Split-middle procedure (White & Haring, 1980) was used to determine the trend lines. Data indicated a rapidly increasing, but variable trend during intervention. Mean frequency of specific social praise during maintenance was 4.4 (range 0-10). Although mean frequency of social praise nearly matched the intervention results during maintenance, a decreasing trend was evident.

Candice. Frequency of Candice's specific praise increased from baseline (M = 0.1, range 0-1) to intervention (M = 2.9, range 1-4). Data indicated a gradually increasing stable trend during intervention. Mean frequency of specific social praise during maintenance was 2.8 (range 0-6). Although mean frequency remained fairly stable from intervention to maintenance, the trend was gradually decreasing.

Susan. Frequency of Susan's specific social praise increased from baseline (M = 0.5, range 0-4) to intervention (M = 4.5, range 3-5). Data indicated a flat trend with high stability, during intervention. During maintenance, mean frequency of specific social praise was 5.8 (range 3-9), which indicated an increase in level from intervention to maintenance. Further inspection of the data during maintenance showed a continually increasing trend as well.

Generalization Probes

Because mentors were not always available during interns' instruction, generalization probes were not collected equally across all conditions. Indeed, some probes were only administered once during baseline and intervention for each intern. Because of this limitation, means are presented along with number of probes in each condition in Table 1.

Tracey. Frequency of Tracey's specific social

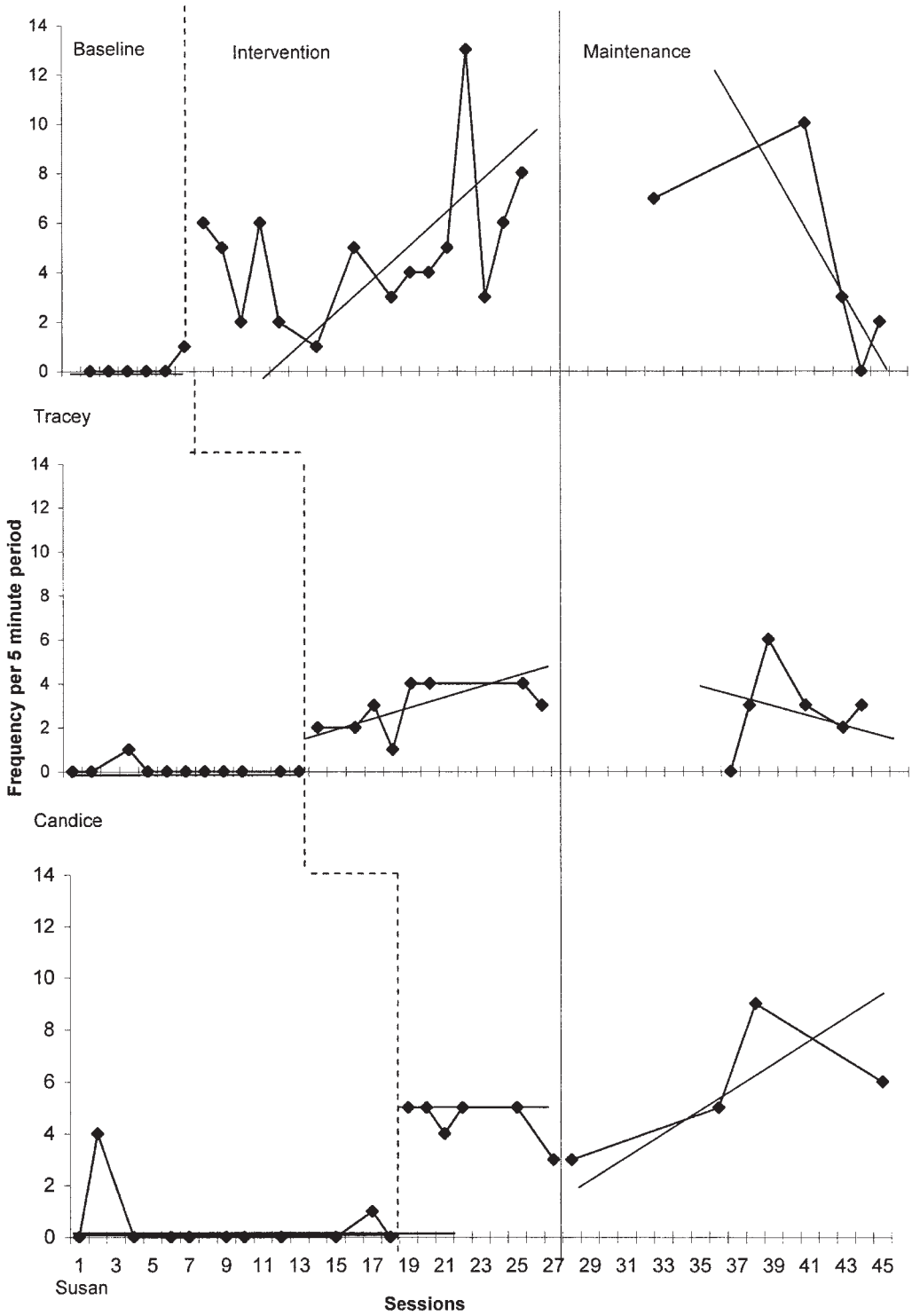

Figure 1. Specific social praise used during teacher-led instruction with trend lines.

TABLE 1

Generalization Probes

Intern	*Condition*	*Mean Frequency of Social Praise (Number of Probes)*	*Content Areas Probed*
Tracey	Baseline	0.0 (1)	Language
	Intervention	0.7 (7)	Reading, Math, Writing, Following Directions Social Studies
	Maintenance	1.4 (5)	Social Studies, Math, Reading
Candice	Baseline	0.8 (4)	Writing
	Intervention	0.0 (1)	Math
	Maintenance	4.4 (5)	Science, Spelling, Writing, Math
Susan	Baseline	0.6 (7)	Spelling, Reading
	Intervention	2.5 (4)	Reading, Spelling
	Maintenance	5.5 (6)	Reading, Social Skills

praise slightly increased from baseline ($M = 0$, range 0) to intervention ($M = 0.7$, range 0-2) during non-targeted contents areas (see Figure 2). However, only one probe was administered during baseline because the mentor was not present during this short baseline to collect data during instruction. Mean frequency of specific social praise during maintenance was 1.4, (range 0-3). Although the level remained fairly stable from intervention to maintenance, the trend during both phases was decreasing.

Candice. Frequency of Candice's specific social praise decreased from baseline ($M = 0.8$, range 0-2) to intervention ($M = 0.0$, range 0) during non-targeted content areas. Only one probe was administered during intervention because the mentor was not present when Candice provided teacher-led instruction during this phase. Mean frequency of social praise during maintenance was 4.4, (range 0-12) and indicated a rapidly increasing, yet variable trend.

Susan. Frequency of Susan's specific social praise increased from baseline ($M = 0.6$, range 0-3) to intervention ($M = 2.5$, range 0-7) during non-targeted content areas and indicated a rapidly increasing, stable trend during intervention. Mean frequency of social praise during maintenance was 5.5, (range 2-12) and indicated an increasing yet variable trend.

Discussion

The purpose of this study was to examine effects of a self evaluation intervention on the instructional behavior of student teacher interns. All three interns immediately increased their use of the targeted behavior (specific social praise) when the intervention was implemented. Social praise use for two interns showed an increasing trend during intervention and the third intern showed stable and substantial increases. Interns also maintained their use of social praise throughout the study.

These findings are consistent with other studies that involved data-based self evaluation procedures to help teachers increase use of effective teaching behaviors (Sutherland & Wehby, 2001; Wright, 1998). These researchers asked classroom teachers to use self evaluation procedures to increase their use of praise statements. Both studies indicated that self evaluation procedures were successful at increasing teachers' use of praise statements.

An additional question was whether any increases would generalize to content areas other than the one targeted. When examining the generalization probes across conditions, two of the three interns did increase their use of specific social praise during non-targeted content areas. The third intern, however, decreased her use of social praise statements during non-targeted content areas. Although results for this question are generally positive they are also inconclusive due to the paucity of data during these conditions. Thus, results of generalization probes must be considered with caution.

Probes during non-targeted content areas in the maintenance condition were administered more consistently and indicated in-

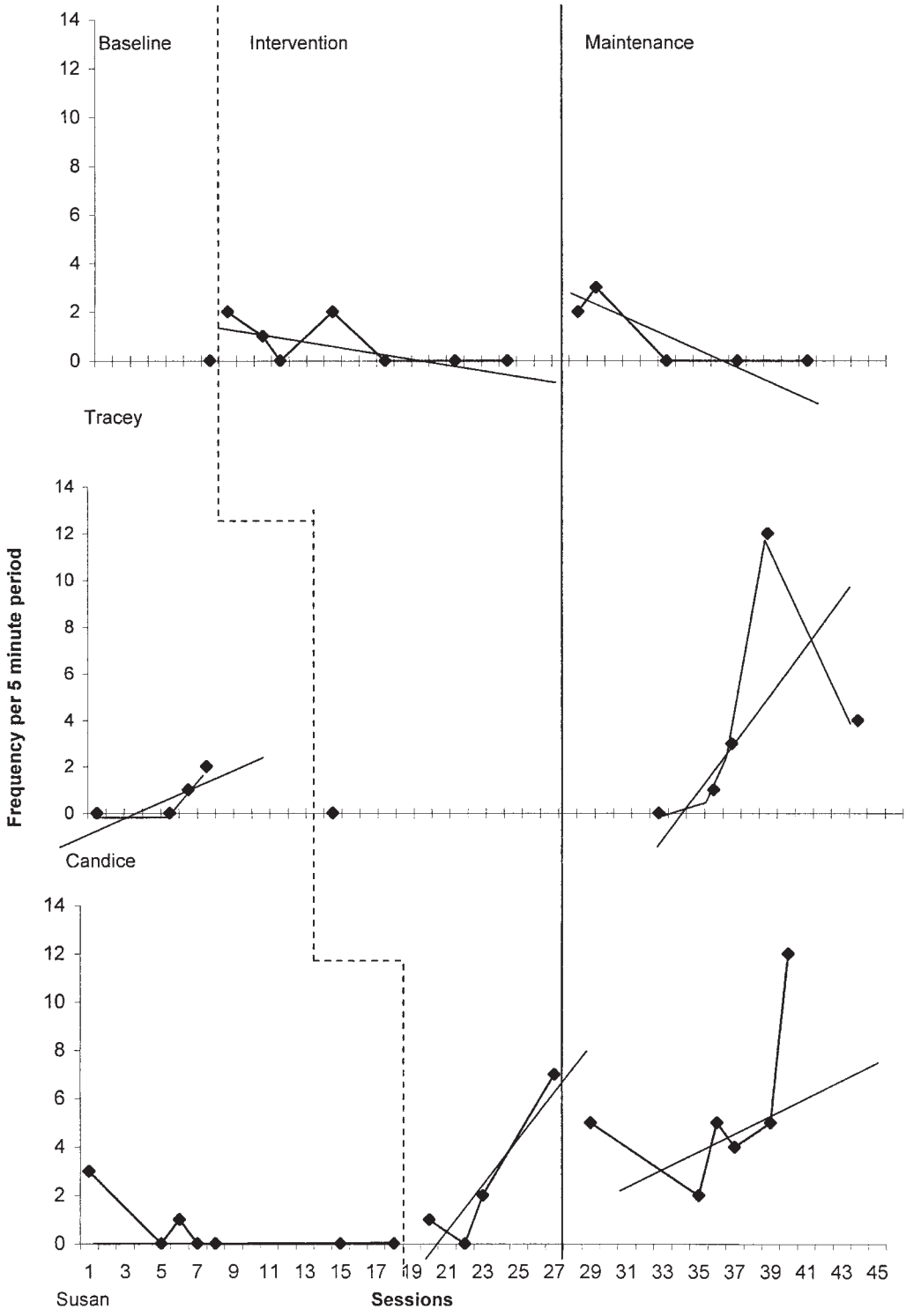

Figure 2. Specific social praise used during non-targeted content areas (generalization) with trend lines.

creased use of social praise statements over baseline.

The final question was to examine whether effects of the intervention would maintain after the intervention was discontinued. All interns' use of social praise statements during maintenance were above baseline. One intern increased social praise statements made from intervention to maintenance while the other two maintained the level reached during intervention.

Although the intervention was successful in maintaining interns' use of social praise for 3½ weeks following the cessation of the intervention, it is possible that additional cues or prompts would have helped them continue to use social praise statements at a more robust level. Also the duration of the intervention may have been too short to build durable responding. Other researchers have shown that longer intervention periods and more gradual dismantling of interventions produce more durable responding when self management interventions are removed (Dougall & Brady, 1998). This assists newer behaviors to develop as a regular part of a pattern of interactions. Unfortunately, time constraints associated with student teaching in this study precluded the opportunity for us to study or strengthen that pattern.

This research adds to the body of literature on data-based self evaluation procedures as a valuable means for promoting increased use of effective instructional practices with teachers (Bullard, 1998; Kilbourn, 1991; Sutherland & Wehby, 2001). However, the results from this study also indicated that a data-based self evaluation strategy is effective for increasing student teacher interns' use of effective practices. These results hold important implications for teacher educators, classroom teachers and teachers in training. For teacher educators, the study found that the self evaluation process was feasible for interns to use to help improve instruction. Teacher educators may want to incorporate a similar strategy into existing student teaching requirements to help interns change their own behavior. This practice could, at a minimum, supplement the external feedback and prompts by supervisors or cooperating teachers used in most teacher education programs. Interns could then learn to examine their own teaching practices, and learn strategies that they can use throughout their teaching careers. As interns become practicing teachers, they may find self evaluation strategies to be effective tools to help them improve their own practice.

In future investigations, researchers should not only look at interns' instructional behaviors, but how teacher self evaluation interventions might affect student social and academic performance. This could be achieved by determining students' task engagement and academic products to determine if the intervention affected academic performance. Additionally, researchers could see if other teacher behaviors (besides social praise) change as a result of intervention. For example, teachers may inadvertently alter the frequency with which they provide academic praise, general praise, and/or teacher reprimands. These and other teacher variables play an important role in students' overall success in school.

References

Airasian, P. W., & Gullickson, A. R. (1997). *Teacher self-evaluation tool kit.* Thousand Oaks, CA: Corwin Press.

Airasian, P. W., Gullickson, A. R., Hahn, L., & Farland, D. (1995). *Teachers' self assessment: The literature in perspective.* Report submitted to the Office of Research, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC.

Alberto, P. A., & Troutman, A. C. (2003). *Applied behavior analysis for teachers.* Upper Saddle River, NJ: Prentice Hall.

Baer, D. M., Wolfe, M. M., & Risley, T. (1968). Current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91–97.

Beaman, R., & Wheldall, K. (2000). The use of approval and disapproval in the classroom. *Educational Psychology, 20,* 431–446.

Brady, M. P., & Taylor, R. D. (1989). Instructional consequences in mainstream middle school classes: Reinforcement and corrections. *Remedial and Special Education, 10*(2), 31–36.

Briggs, L. D., Richardson, W. D., & Sefzik, W. P. (1986). Comparing supervising teacher ratings and student teacher self-ratings of elementary student teaching. *Education, 106,* 150–159.

Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*(1), 5–32.

Buchanan, D., & Jackson, S. (1998). Supporting self-evaluation in initial teacher education. *Scottish Educational Review, 30,* 110–124.

Bullard, B. (1998, November). *Teacher self-evaluation.*

Paper presented at the annual meeting of the Mid-South Educational Research Association. (ERIC Document Reproduction Service No. ED428074)

Emmer, E. T. (1988). Praise and the instructional process. *Journal of Classroom Interaction, 23,* 32–39.

Englert, C. S. (1984). Effective direct instruction practices in special education settings. *Remedial and Special Education, 5,* 38–47.

Freiberg, H. J., & Waxman, H. C. (1990). Reflection and the acquisition of technical teaching skills. In R. Clift, W. Houston, & M. Pugach (Eds.), *Encouraging reflective practice in education* (pp. 119–137). New York: Teachers College Press.

Gable, R. A., Hendrickson, J. M., Young, C. C., Shores, R. E., & Stowitschek, J. J. (1983). A comparison of teacher approval and disapproval statements across categories of exceptionality. *Journal of Special Education Technology, 6,* 15–22.

Greis, N. (1986). Self evaluation in practice teaching. *Foreign Language Annals, 19,* 231–236.

Griffin, C. C., & Kilgore, K. L. (1995). Framing the problem of practice: The effects of self-assessment in a study of special education students' internships. *Teacher Education and Special Education, 18,* 56–71.

Gunter, P. L., & Coutinho, M. J. (1997). Negative reinforcement in classrooms: What we're beginning to learn. *Teacher Education and Special Education, 20,* 249–264.

Gunter, P. L., & Reed, T. M. (1996). Self-evaluation of instruction: A protocol for functional assessment of teaching behavior. *Intervention in School and Clinic, 31,* 225–230.

Guyton, E., & McIntyre, D. J. (1990). Student teaching and school experiences. In W. R.

Houston (Ed.), *Handbook of research on teacher education* (pp. 514–534). New York: Macmillan.

Hancock, D. R. (2000). Impact of verbal praise on college students' time spent on homework. *Journal of Educational Research, 93,* 384–389.

Kagan, D. M., & Tippens, D. J. (1991). Helping interns attend to student cues. *The Elementary School Journal, The Elementary School 91,* 343–356.

Kilbourn, B. (1991). Self-monitoring in teaching. *American Education Research Journal, 28,* 721–736.

McDougall, D., & Brady, M. P. (1998). Initiating and fading self management interventions to increase math fluency in general education classes. *Exceptional Children, 64,* 151–166.

Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, S. B. (1999). *Single subject research: Applications in educational and clinical settings.* San Diego: Singular.

Speer, W., & Zoellick, C. (1974). *Self-evaluation: Developing the skill.* Prepared under the sponsorship of the Academic Affairs Conference of Midwestern Universities and the Chicago Consortium of Colleges and Universities. (ERIC Document Reproduction Service No. ED249182)

Stronge, J. H. (1997). Evaluating teaching: A guide to current thinking and best practice. *Thousand Oaks, CA: Corwin Press.*

Sutherland, K. S. (2000). Effects of self-evaluation on rates of teacher behaviors in classrooms for student with emotional and behavioral disorders. *Dissertation Abstracts International, 61,* (11), 4340 (UMI No. 9996222).

Sutherland, K. S., & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms for students with emotional and behavioral disorders. *Journal of Special Education, 35,* 161–171.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8,* 2–8.

Sutherland, K. S., Wehby, J. H., & Yoder, P. J. (2002). Examination of the relationship between teacher praise and opportunities for students with EBD to respond to academic requests. *Journal of Emotional and Behavioral Disorders, 10,* 5–13.

Van Acker, R., Grant, S. H., & Henry, D. (1996). Teacher and student behavior as a function of risk for aggression. *Education and Treatment of Children, 19,* 316–334.

Wehby, J. H., Symons, F. J., & Shores, R. E. (1995). A descriptive analysis of aggressive behavior in classrooms for children with emotional and behavioral disorders. *Behavioral Disorders, 20,* 87–105.

Wheeler, A. E., & Knoop, H. R. (1982). Self, teacher, and faculty assessments of student teacher education. *Journal of Educational Research, 75,* 178–191.

White, O., & Haring, N. (1980). *Exceptional teaching* (2nd ed.). Columbus, OH: Charles E. Merrill.

Wright, M. R. (1998). *The effects of video-based teacher self-evaluation on teacher instructional performance.* Unpublished doctoral dissertation, University of South Alabama.

Received: 10 March 2004
Initial Acceptance: 26 April 2004
Final Acceptance: 15 July 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 377–389

Selecting and Validating Tasks from a Kindergarten Screening Battery That Best Predict Third Grade Educational Placement

Marcia Strong Scott, Christine F. Delgado, Shihfen Tu and Kathryn L. Fletcher
University of Miami

Abstract: In this study, predictive classification accuracy was used to select those tasks from a kindergarten screening battery that best identified children who, three years later, were classified as educable mentally handicapped or as having a specific learning disability. A subset of measures enabled correct classification of 91% of the children in those special education categories and 85% of the children in regular education. Although there was a high percentage of false positives, 57% of these children were actually poor achievers in third grade. Lower screening scores were associated with being tested in Spanish or both English and Spanish, minority status and low SES.

More than a decade ago, Public Law 99-457 (1986) mandated the provision of public school education for preschool children with handicaps. This legislative act necessitated identification of all preschool children eligible for publicly funded early specialized educational programs. Of all those who are eligible, the group most difficult to identify at an early age are those with mild learning problems (Jacob, Snider, & Wilson, 1988; Meisels, 1989; Mercer, Algozzine, & Trifiletti, 1988; van Kraayenoord, 1983). They are identified later in grade school, even when an extensive screening program is in place (Scott, Urbano, & Boussey, 1991). Children classified as educable mentally handicapped (EMH) or as having a learning disability (LD) are examples of two categories of special needs children typically described as having mild learning problems (Gresham, MacMillan, & Bocian, 1996; Holtzman & Wilkinson, 1991).

Clearly, full implementation of PL 99-457, now called the Individuals With Disabilities Education Act, will not be possible until screening tests are developed that more effectively identify young children with mild learning problems. Use of such screening measures presupposes that the ". . . behavior and cognition of young children is malleable and that early intervention will facilitate appropriate development" (Salvia & Ysseldyke, 1991, p. 467). Consistent with this belief, early intervention has proven to be a) more effective for those who participated from an early age (Shonkoff & Hauser-Cram, 1987), b) associated with a reduced need for subsequent special education (Lazar & Darlington, 1982), and c) more effective for children with mild delays (Dunst, Trivette, & Cross, 1986).

As one potential remedy for the difficulties encountered in the early detection of children with mild learning problems, Scott and her colleagues (Scott, Greenfield, & Sterental, 1986) have been designing and evaluating new cognitive screening tests that are composed of simple tasks, but tasks that require the active engagement of children's cognitive processing skills (Greenfield & Scott, 1985).

Shihfen Tu is now at the University of Maine. Kathryn Fletcher is now at Ball State University. The authors would like to thank all the administrative and teaching staff that aided us in the execution of this study. This research was supported in part by a grant from the National Institute of Child Health and Human Development (HD22952) and by the Florida Diagnostic and Learning Resources System/ University of Miami, a specialized university center funded through State General Revenue appropriations to provide multidisciplinary evaluation services in exceptional student education programs. Correspondence concerning this article should be addressed to Marcia S. Scott, Department of Psychology, Annex, P.O. Box 248185, Coral Gables, FL 33124-0751. E-mail: mscott@miami.edu

For example, children might be shown a picture consisting of many cats and be asked to tell what they look like and what they do (Scott, Fletcher, Jean-Francois, Urbano, & Sanchez, 1998) or they might be shown several arrays of three, nine-square matrices, two identical and one different pattern of colored dots displayed on each three-matrix array, and be required to point to the matrix with the different design (Scott, Deuel, Urbano, Fletcher, & Torres, 1998). To respond correctly, children must activate their cognitive processing capabilities and analyze the stimulus problem. A simple automatic associative response will not do. The use of their cognitive processing capabilities was expected to make overt, their not-yet-identified mild cognitive deficiencies.

Use of cognitive tasks is consistent with the finding that cognition is one of the developmental areas related to subsequent educational status (Lichtenstein & Ireton, 1991) and later school achievement (Funk, Sturner, & Green, 1986). Additionally, cognitive capabilities assessed prior to, or at the beginning of kindergarten have been shown to relate to both reading and math performance (e.g., Mantzicopoulos & Morrison, 1994; Swanson, 1994; Tramontana, Hooper, & Selzer, 1988). These are major areas of learning difficulty for children classified as EMH or LD, the latter condition now called specific learning disability (SLD), which more clearly reflects the limited extent of the problems typically observed with these children.

Following evaluations of individual tasks (e.g., Scott, Perou, Greenfield, & Swanson, 1993) and then batteries of multiple tasks (e.g., Scott, Deuel, Claussen, & Sanchez, 1993), a series of three concurrent validity studies was conducted with preschool children in which several versions of a screening battery were evaluated (e.g., Scott, Deuel, et al., 1998). For each version, the accuracy with which preschool children already classified as EMH or SLD and children without learning problems could be differentiated was investigated. Classification accuracy is the recommended method for assessing any screening test (e.g., Lichtenstein, 1981; Lichtenstein & Ireton, 1984, 1991). In the final concurrent evaluation (Scott, Fletcher, Jean-Francois, et al., 1998), classification accuracy was determined separately for the preschool and kindergarten samples based on results from another study (Scott, Fletcher, & Deuel, 1998) and it was determined that classification accuracy in this final concurrent evaluation study was also maximized with different subsets of tasks for the preschool and kindergarten samples.

Subsequent to this study, separate evaluations were conducted of the preschool screening set and the kindergarten screening set. The preschool screening set was cross validated (Scott, Tu, & Fletcher, 2003) and test-retest reliability was demonstrated (Jean-Francois, Urbano, & Scott, 1999). The kindergarten set, however, did not demonstrate acceptable levels of reliability.

While concurrent validity provides necessary information about any new test, one must also assess the long-term prediction of a test (Rafoth, 1997) so that examiners can judge how well initial screening results can predict future diagnostic/placement decisions (Lenkarski, Singer, Peters, & McIntosh, 2001). In the 1993/94 school year, the initial version of the screening battery was administered to a large sample of normally achieving preschool and kindergarten children. These samples were followed into public school through third grade. Their third grade educational status was available. Linking the children's screening scores to their third grade educational status would show whether a brief screening measure administered to preschool and kindergarten children could accurately identify which children were at risk for an EMH or SLD label in grade school.

We first examined the predictive capabilities of the screening performance of the preschool sample. The results were encouraging. We found that four years later, a set of just three tasks from the battery accurately predicted the third grade educational placement of 79% of the children with an EMH or SLD label and 70% of the children who were in regular education (Scott & Delgado, 2003). In this study, predictive accuracy of the screening performance of the kindergarten sample was evaluated. Once again, their educational status in third grade was the primary outcome measure.

Given previous results in both concurrent (e.g., Scott, Deuel, et al., 1998; Scott, Fletcher,

Jean-Francois, et al., 1998) and predictive (Scott & Delgado, 2003) studies, as well as data from other researchers (e.g., McIntosh, 1999; Stebbins & McIntosh, 1996), it was expected that the most accurate predictions would be associated with a subset of tasks from the screening battery, not with the total screening score.

Minority status (Scott et al., 2003; Scott, Fletcher, Jean-Francois, et al., 1998) and testing in Spanish or both Spanish and English (Scott et al., 2003) were associated with lower screening scores. In the preschool predictive study (Scott & Delgado, 2003), testing in Spanish or both languages was shown to be confounded with low SES, which in turn was related to poor screening performance and risk for poor long-term academic outcome. The confounding of minority status with low SES is consistent with other findings (e.g., Halle, Kurtz-Costes, & Mahoney, 1997) as is the relation between poor academic performance and low SES (Hill, 2001). The relation of testing language, minority status and SES to screening performance and academic achievement were further explored in the present study.

Method

Participants

Kindergarten sample. Four hundred and fifty-nine kindergarten children were administered a screening battery consisting of nine cognitive tasks in the 1993-94 school year. This sample was composed of every child from the targeted kindergarten classes in the 35 participating public schools for whom an informed consent was obtained. Informed consent included permission to access the child's educational status and SAT scores through third grade. All six regions of the Miami/Dade County Public School (M/DCPS) system were represented. According to parental designations, 39% of the sample were classified as Black, 34.5% as Hispanic, and 26.5% as White. These three racial/ethnic classifications were the only ones used by the school system at that time. Clearly minority children make up most (73.5%) of the kindergarten sample. Fifty-nine percent of the sample was male and the mean chronological age was 72.0 months (SD = 4.2).

Third grade predictive sample. Of the 459 kindergarten children with screening scores, 356 were successfully tracked into third grade. Of these, 345 were in regular education and 11 were in special education with either an EMH (1) or SLD (10) classification.

According to statewide criteria, for a child to be classified as EMH, he or she must achieve a full scale IQ score that falls between 2 and 3 *SD*s below the mean on two measures of intelligence and achieve a score on a test of adaptive behavior such as the Scales of Independent Behavior-Revised (Bruininks, Woodcock, Weatherman, & Hill, 1996) or the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) that falls at least 1.5 *SD*s below the mean of a child of the same age and socio-cultural group.

According to the same state wide criteria, for a child to be classified as LD, there must be a) a significant difference between language performance and other developmental levels, or b) a significant difference between receptive and expressive language abilities, or c) a significant language delay based on criteria presented in the test or evaluation manual.

All six regions in the M/DCPS system were still represented in the third grade predictive sample. The racial/ethnic distribution of this smaller sample was 41% Black, 35% Hispanic and 24% White. Fifty-six percent were male. Mean chronological age of this sample when the children were in kindergarten was 72.0 (SD = 4.3). This primarily minority sample (76%) is nearly identical to the kindergarten sample in terms of race/ethnicity and also mimicked the total school population which was approximately 83% minority in the 93/94 school year, the year the kindergarten children were administered the screening battery.

Kindergarten children excluded from the predictive sample. There were 101 children not included in the predictive sample. Seven were excluded because they were in a grade other than third, the grade they should have been in during the 1996-97 school year, and the reasons for this were not known. Twenty-seven children were never found in the school records after kindergarten and 57 children were not present in the third grade records although they had been found in earlier

grades. An additional 12 children were excluded because they were in special education classifications other than EMH or SLD.

There were more White children (35%), fewer Black children (33%), and more males (67%) in the group of children not included in the predictive sample. However, the mean chronological age of this excluded group was 72.0 (SD = 3.9) which was the same as mean chronological age of the 356 children in the predictive sample.

Testing Language

If a child lived in a home in which a language other than English was spoken, that child was tested in the language in which the child was most proficient. This was determined through a consideration of teacher nomination, the child's choice and the child's conversational ability. Spanish was the only other language used for testing. If the examiner was not certain that the child understood the instructions in the language employed, the examiner administered instructions in both languages. Of the 125 children in the third grade predictive sample designated Hispanic, 81% chose to be tested in English. Fifteen percent chose to be tested in Spanish and the screening battery was administered in both languages to only 4%. Examiners fluent in Spanish translated instructions into that language for those children who needed to be tested in Spanish or both languages. Responses in Spanish were accepted from any child regardless of testing language.

Screening Battery Contents and Administration

The screening battery consisted of nine tasks. Test items included colored photographs of meaningful pictures, or in one case colored dots. These stimuli were placed on 35.6 cm wide by 21.6 cm high white paper pages, which were placed in a four-hole legal size black binder, 36.8 cm wide by 25.4 cm high. The tasks are briefly described below in the same order as they appeared in the screening battery. For more details about the tasks and scoring see Scott, Deuel, et al. (1998).

Picture pointing. Six pictures were displayed on each of four pages. Children were asked to point to each picture just one time. The score for each array ranged from 0 (omission or repetition present) to 3 (pointing in a reading order sequence). Two dependent measures were evaluated: a quality score, which was the sum of the scores over the four pages, and a penalty score which was the quality score minus the total number of omissions and repetitions.

Picture recognition. After presenting a training memory page and test page, children were shown two sets. Each set consisted of one memory page and two recognition test pages. The examiner named each of the eight pictures on the memory page twice. On each of the two recognition test pages there were four "seen before" pictures and four new pictures. As test pages were presented, children were asked to point to the pictures that were the same as those from the memory page. The dependent measure was number of correct selections minus number of incorrect selections summed over the two sets.

Word meaning. Children were asked, "What is an airplane?" and then "What is a banana?" Scores for each part of the definition varied from 0 (statement not true) to 3 (abstract categorical).

Standard oddity. Children were asked to point to the different picture in each of nine unique arrays consisting of two identical and one different picture. The dependent measure was number of correct selections out of nine.

Dot matrix oddity. On each of seven pages, there were three 7.6 X 7.6 cm matrices. Each matrix was divided into nine squares. Two of the matrices had an equal number of dots placed in exactly the same position while on the different matrix, the dots were in the same positions but there was either one more or one less dot. Children were asked to point to the different design, the one not like the others. The dependent measure was number of correct selections out of seven.

Sequencing. On each of four pages, there was a sequence of either colored dots or animals running across the upper half of the page ending in a missing item. Children were asked to point to the one of three colored dots or animals, located on the bottom of the page that was missing from the top. The dependent measure was number of correct selections out of four.

Picture rhyming. Children were asked to point to the pair of pictures whose names rhymed, sounded alike. One pair appeared above a horizontal black line and the other below it. There was one training page and two test pages. The dependent measure was number correct out of two. Regardless of the language used to administer the instructions, the names rhymed in English, and were labeled in English.

Unstructured semantic information. Children were first shown a collage of 15 cats and then a grouping of fruit. They were asked to tell all they knew about cats and fruit. One point was awarded for each description that was valid for all, or most exemplars. The dependent measures were total number of points awarded to cats and to fruit.

Structured information. Pictures of four different people, who varied in race/ethnicity, age and gender, were displayed on a single page. Children were asked to tell how people differed from one another. One point was awarded for each valid difference. The dependent measure was total number of points awarded.

Testing Procedure

All children were tested individually, for a single session, by one of six female examiners. Before starting the testing session, examiners pointed out a selection of award certificates and stickers that the children were told they could earn by playing the game. Regardless of their performance, all children were given an award certificate and stickers. Presentation time was typically about 15 to 20 minutes. During presentation of the test, children were periodically praised for their good performance.

Results

Frequency distributions showing number of children in regular education and in special education who achieved each score were computed. At each cut score, all children with screening scores below the cut were said to be at risk for a special education placement, while those with scores at, or above the cut were predicted to be in regular education in third grade. Classification accuracy was determined by comparing the predicted to the actual educational placement of each child in third grade. The cut score selected was the one associated with the best combination of sensitivity and specificity (Yerushalmy, 1947). In this study, sensitivity was the percentage of all children with an EMH or SLD classification in third grade whose scores fell below the cut and specificity was defined as the percentage of all children in regular education in third grade whose scores were at, or above the cut.

Classification Accuracy Using the Total Screening Score

Since there were two measures to be evaluated for the picture pointing task, one frequency distribution used the quality score to represent the contribution of this task to the total score and the second distribution used the penalty measure. The best combination of sensitivity (82%) and specificity (79%) was achieved at a cut score of 60, using the penalty measure to represent the contribution of the picture pointing task to the total screening score. These values became the base against which estimates of sensitivity and specificity using subsets of tasks were compared.

Classification Accuracy Using Subsets

First, the best combination of sensitivity and specificity achievable for each individual task was determined. Since there were two measures for the picture pointing, unstructured semantic information and word meaning tasks, 12 measures were examined in this initial evaluation. The best combination of sensitivity (82%) and specificity (68%) was associated with scores from the structured information task.

Each child's score on the structured information task was then combined with his/her score on each of the other measures, which resulted in the computation of 11 new frequency distributions, each using a score summed over two tasks. The best two-task set was chosen and additional measures were added (i.e., best three-task combination, four-task combination, etc.) until no further improvement in either sensitivity and/or specificity was obtained. In this instance, the highest levels of sensitivity (91%) and specific-

ity (85%) were achieved using a five-task set consisting of the structured information, standard oddity, word meaning/banana, dot matrix oddity and unstructured semantic information/fruit tasks, with a cut score of 21. The distribution for this subset is shown in Table 1. Scores have been blocked in order to illustrate the findings in a single table. This combination improved on the accuracy levels obtained with the full battery in terms of both sensitivity (91% vs. 82%) and specificity (85% vs. 79%).

Other Psychometric Characteristics

As well as evaluating the sensitivity and specificity one can achieve with a screening test, it is also important to know, for example, the percentage of underreferrals associated with the test. This measure was defined as the percentage of children with an SLD or EMH classification whose screening scores were at, or above the cut score (see Table 1). Only one of the eleven children (9%) with an SLD or EMH classification earned a score that was not below the cut and was, therefore, incorrectly predicted to be in regular education in third grade. In addition, only 15% of the children in regular education in third grade had kindergarten screening scores below the cut. These students are called false positives because they were incorrectly predicted to be in special education in third grade. Both of these values represent good psychometric characteristics.

However, predicting three years into the future led to an 84% over referral rate. That is, of all kindergarten children with scores below the cut who were predicted to be at risk for a special education placement, 84% were not in special education in third grade. Only 16% of children with scores below the cut actually did have an EMH or SLD label by third grade. This value represents the positive predictive accuracy of the set (Feinstein, 1976) and, on the face of it, is too high.

Over Referral and Poor Achievement

Percentile achievement scores on the reading comprehension, math computation and math application subtests of the Stanford Achievement Test (SAT) were available for most of the children. In the preschool predictive study (Scott & Delgado, 2003), many children in regular education who had preschool screening scores below the cut (false positives), ac-

TABLE 1

Number of Children in Regular Education and Special Education Achieving Each Score on the Five-Task Subset That Included the Structured Information, Standard Oddity, Word Meaning/Banana, Dot Matrix Oddity, and Unstructured Semantic Information/Fruit Tasks

Score	*Regular Education*	*Special Education*
8–10	2	0
11–13	0	1
14–16	11 ($n = 52$)	1 ($n = 10$)
17–19	22	3
20	17	5
	Cut Score	
21–23	70	0
24–26	80	0
27–29	75	1
30–32	39 ($n = 293$)	0 ($n = 1$)
33–35	21	0
36–38	3	0
40–42	3	0
44–49	2	0
	345	11
SENSITIVITY = 10 OF 11 OR 92%		SPECIFICITY = 293 OF 345 OR 85%

tually did demonstrate learning difficulties in third grade. They earned very low achievement scores on one or more of the three achievement tests available in the data set. For the purpose of examining this aspect of academic performance, we defined poor achievement as having one or more achievement scores at, or below the 15th percentile. Children excluded were those with no SAT data and those with one or two missing scores whose recorded score(s) was greater than the 15th percentile. For example a child with one score at the 25th percentile would be excluded because we did not know what scores would have been achieved on the other two subtests; one or more might have been at or below the 15th percentile. However, if a single score was at or below the 15th percentile, even if one or more subtest scores were missing, then that child was classified as a poor achiever. Only children who could be classified as either a poor achiever or not were included in analyses of achievement data.

The same definition of poor achievement was used in this study. Of the 345 children in regular education, 340 could be classified as poor achievers or not. Overall, 79 or 23% were poor achievers in third grade. Of more importance to an evaluation of the over referral problem, 49 of the 52 children in regular education with screening scores of less than 21 could be classified. Of these 49, 28 or 57% were poor achievers in third grade. This represented approximately 35% of all children who were poor achievers in third grade.

If we now look at the predictive accuracy associated with predicting risk for learning problems in third grade, where learning problems are defined as either showing poor achievement or having an EMH or SLD label, then the positive predictive accuracy associated with the five-task set over three years is 64%, not 16%. This numerator for this computation was 38 which was the sum of the number of children with an EMH or SLD label ($n = 10$) plus the number of regular education children with scores below the cut who were poor achievers ($n = 28$). The denominator was 59, which was the sum of the total number of children in regular education with scores below the cut who could be classified ($n = 49$) plus the number of children with an EMH or SLD label with a score of less than 21 ($n = 10$). The percentage just drops to 61 if all children in regular education with scores below the cut are used, rather than only those who could be classified as a poor achiever or not.

Variables Impacting Kindergarten Screening Performance

In all of the following analyses, children's screening scores on the five-task subset were used because this was the subset that maximized predictive accuracy. Unless otherwise indicated, the entire predictive sample of 356 were included because in kindergarten, when the children actually were administered the screening battery, none was in special education. However, all analyses of achievement performance included only the normally achieving group from the predictive sample because with this measure we were looking for children with poor achievement who were in regular, not special education.

Gender. Although males ($M = 24.8$; $SD = 5.3$) had significantly lower screening scores, $F(1, 354) = 6.36$, $p < .02$, than females ($M = 26.2$; $SD = 5.2$), the higher percentage of male (20.5%) than female (13%) children with scores below the cut was not significant, X^2 (1, $N = 356$) = 3.02, $p > .05$, nor was the difference between males (64%) and females (44%) in terms of the percentage of children with scores below the cut who were poor achievers in third grade, X^2 (1, $N = 49$) = 1.74, $p > .05$.

Testing language. We first compared the Hispanic/English sample ($n = 101$) with the Hispanic/Spanish sample ($n = 19$). The Hispanic/Spanish group ($M = 20.9$, $SD = 5.8$) had significantly poorer screening scores, $F(1, 118) = 12.58$, $p < .01$, than the Hispanic/English group ($M = 25.5$, $SD = 5.1$). There was also a significantly higher proportion of children from the Hispanic/Spanish group (37%) than from the Hispanic/English group (17%) who earned scores of less than 21, X^2 (1, $N = 120$) = 4.00, $p < .05$. These differences favoring Hispanic children tested in English, however, did not impact actual placement as all five Hispanic children in special education in third grade were tested in English.

Both the results pertaining to the screening

score measure, $F(1, 123) = 12.35$, $p < .01$, and those reflecting the proportion of children with scores below the cut, X^2 (1, $N = 125$) = 5.03, $p < .02$, were replicated, when the Hispanic/Spanish group was combined with the Hispanic/both group ($n = 24$) to form the Hispanic/Spanish-both group ($M = 21.4$, $SD = 5.6$).

Testing language did, however, impact achievement status. The combined Hispanic/Spanish-both group was used ($n = 24$) to increase the sample size for this categorical test. Considering only those children with scores below the cut, 23% of those tested in English were poor achievers compared to 78% of those children tested in Spanish or both languages. The difference in proportions was significant, X^2 (1, $N = 22$) = 6.42, $p < .03$.

Race/ethnicity. Children in the White sample ($M = 27.7$, $SD = 3.8$) earned significantly higher screening scores, $F(2, 353) = 10.74$, $p < .001$, than children in the Hispanic ($M = 24.7$, $SD = 5.5$) and Black ($M = 24.6$, $SD = 5.5$) samples, both p values $< .001$, while the two minority groups did not differ significantly from each other, $p = .86$.

Children in the Black and Hispanic samples were also nearly six times more likely to earn a score below the cut, 23% and 21% respectively, than were children from the White sample (4%). For the race/ethnicity variable, the differential risk for a special education placement was validated. Of the 11 children classified as either SLD or EMH, 9% were White, but 45.5% were Black and 45.5% were Hispanic.

With respect to poor achievement, two children from the White sample had scores below the cut and one of these was a poor achiever in third grade. Such a small sample is unlikely to produce a reliable estimate. Consequently, a comparison across the three groups was not feasible. However, 68% of Black children and 45% of Hispanic children, with scores less than 21 were poor achievers in third grade.

The impact of testing language on race/ethnicity. Did the Hispanic sample as a whole perform at a lower level than the White sample or was it only those Hispanic children who were tested in Spanish or both Spanish and English? To answer this question, the Hispanic sample was divided into two groups: Hispanic/English ($n = 101$) and Hispanic/Spanish ($n = 19$). The other two racial/ethnic groups were also delineated in terms of their testing language: White/English ($n = 85$) and Black/English ($n = 146$).

An analysis of variance (ANOVA) of groups was significant, $F(3, 347) = 11.71$, $p < .001$. Post hoc tests showed that while the screening scores of the White/English group were still significantly higher than the scores of all three minority groups, all p values $< .01$, both the Hispanic/English ($M = 25.5$; $SD = 5.1$) and Black/English ($M = 24.6$; $SD = 5.5$) groups had significantly higher screening scores than the Hispanic/Spanish group ($M = 20.9$; $SD = 5.8$), both ps $< .01$. The same results were observed when children tested in both English and Spanish were added to the Hispanic/Spanish group.

A similar pattern of mean differences was observed when the percentage of children with scores below the cut was examined. The White/English sample had the fewest (4%), the Hispanic/English and Black/English had percentages in the middle, 17% and 23% respectively, and 37% of the Hispanic/Spanish group earned scores below the cut. There was a similar percentage of children with scores below the cut (38%) for the Hispanic/Spanish-both group. However, remember that of the five Hispanic students in special education, all were tested in English.

A slightly different pattern of group differences was observed when the dependent measure was the percentage of children in regular education, with scores below the cut, who were poor achievers in third grade. Only two children from the White/English sample had scores below the cut, so no reasonable estimate can be made of the percentage of poor achievers for this group. However, only 23% of children from the Hispanic/English group with scores below the cut were poor achievers while 68% of children in the Black/English group and 78% of children in the Hispanic/Spanish-both group with scores below the cut were poor achievers in third grade.

SES. SES of each child was estimated using the percentage of children eligible for free or reduced lunch in his/her school. A child was eligible for reduced lunch if annual household income was at, or less than, 185% of the poverty level, which is defined annually by the federal government. A child was eligible for

free lunch if annual household income was 130% of the poverty level. Since eligibility is in terms of household income level, the percentage of children in each school who are eligible for these services is a good estimate of the SES of the families in that catchment area.

Schools were divided into three SES groups; low, mid-level and high. Schools were designated as serving a low SES population if from 70.6% to 99% of students in that school were eligible for free or reduced lunch. Schools were placed in the mid-level SES group if from 30.6% to 70.5% of their population were eligible for free or reduced lunch. Schools with less than 30.6% of their population eligible for free or reduced lunch were designated high SES. There were 15, 15 and 3 schools in the low, mid-level and high SES conditions, respectively.

A one-way ANOVA of SES group was computed with scores on the five-task subset as the dependent measure. Mean screening score increased as SES increased from low (24.2; SD = 5.6), through mid-level (25.9; SD = 5.4) to high (26.7; SD = 3.8). These differences were significant, $F(2,353) = 5.89$, $p < .01$. Post hoc tests showed that both the high and mid-level SES condition were associated with higher screening scores than the low SES condition, both p values $< .01$, but were not significantly different from each other, $p = .29$.

The percentage of children with screening scores below the cut increased as SES level decreased. Only five percent of children in the high SES group, but 14% of children in the mid-level group and 21% of children attending low SES schools earned such low kindergarten screening scores. Similarly, 9% of the children in special education were in a high SES school in kindergarten compared to 45.5% who were in mid-level SES schools and 45.5% who were in low SES schools.

A similar relation was present in terms of the percentage of children with kindergarten screening scores below the cut who were poor achievers in third grade. Again, only three children from the high SES condition had scores below the cut and one of these was a poor achiever in third grade; a small sample to estimate from. However, 57% (12 of 21) of children with scores below the cut in the mid-level SES condition, and 60% (15 of 25) of children with scores below the cut in the low SES condition were poor achievers in third grade.

Impact of SES on race/ethnicity. Since there were so few Hispanic children tested in Spanish ($n = 19$) or both languages ($n = 5$), these two groups were combined when the impact of SES on screening performance was examined. Number of children in the White/English, Hispanic/English, Black/English and Hispanic/Spanish-both groups who were in each of the three SES conditions was examined. There were eight or fewer students in five cells of the twelve-cell matrix. This precluded using an ANOVA. However several relationships between SES and other variables were observed.

With respect to race, the White/English sample had the highest percentage (40%) of students in high SES schools while not one student in the Black/English group attended a high SES school. In contrast, only 5% of the White sample but 77% of the Black/English sample were in low SES schools.

With respect to ethnicity and testing language, more children from the Hispanic/English (24%) than from the Hispanic/Spanish-both sample (12.5%) were attending high SES schools, while fewer children in the Hispanic/English sample (10%) than in the Hispanic/Spanish-both sample (29%) were enrolled in low SES schools.

Discussion

The major intent of this study was to evaluate a kindergarten screening battery, and selected items from that battery, in terms of the accuracy achieved when kindergarten screening scores were used to predict third grade educational status.

Children's total scores on the kindergarten screening battery were associated with a sensitivity of 82% and a specificity of 79%. However, consistent with previous data (e.g. Scott, Deuel, et al., 1998, Stebbins & McIntosh, 1996), a higher level of both sensitivity (91%) and specificity (85%) was achieved with a five-task subset consisting of the structured information, standard oddity, word meaning/banana, dot matrix oddity and unstructured semantic information/fruit tasks. The inclusion of three generating tasks in this five-task subset is consistent with the position advo-

cated by Simmons and Kameenui (1990) that performance differences between intact and mildly impaired groups will be greater when those groups are compared using tasks that a) require the participants to produce responses and b) have little external support. Generating tasks fit this description.

Both the sensitivity and specificity associated with this five-task set are above the minimum 80% levels recommended by Meisels (cited in Lichtenstein & Ireton, 1991). They are also exceptionally high when considering the influence on school performance of potentially huge variations in the home and school environments the children were likely to have experienced during the three years between the kindergarten assessment and third grade. Varied experiences can be expected to differentially affect performance and contribute to the difficulty of achieving accurate long term prediction (Lichtenstein & Ireton). Indeed, these authors suggested that with preschool screening, perhaps one must settle for a "reasonable correspondence with a concurrent developmental assessment" (p.503). Results of this study show, however, that early and accurate identification of kindergarten children who will have mild learning problems later in school is possible.

Although nearly all children with a special education placement of either SLD or EMH were accurately identified, there was a very high level of over referral. A high rate of over referral is often found with developmental screening measures (Gredler, 1997; Lichtenstein & Ireton, 1984) and may be a necessary consequence of trying to identify a low prevalence group. Another factor bearing on the actual impact of the high percentage of over referrals to the utility of the screening set is the finding that of the children in regular education with scores below the cut whose achievement status could be ascertained, 57% were poor achievers. In short, a high percentage of false positives using the placement-in-special education criterion were actually evidencing poor achievement, another type of learning problem. These results replicate those reported in Scott & Delgado (2003) for a preschool predictive sample where there was also a high percentage of false positives who were poor achievers.

If one considers all those children with scores below the cut of 21 who were either in special education or were poor achievers, then the positive predictive accuracy associated with the identification of children at risk for either poor achievement or a special education placement is 64%, a value much higher than the 16% associated with risk for special education placement alone. Indeed, if the most critical aspect of a screening test is its ability to identify those children who will have problems in school (e.g., Limbos & Geva, 2001; Rafoth, 1997), the five-task subset identified in this study can be considered to be a successful and effective screening device.

Since 36% of kindergarten children with scores below the cut were neither in special education nor were poor achievers in third grade, one is hesitant to refer all children with scores below the cut for a costly psychoeducational assessment. Rather, results of the children's performance on this very brief screen could be used to identify a subsample of children who are at risk for school problems, a group who should be carefully monitored for possible learning problems. Then, at the earliest sign of any academic problems by anyone from this risk group, the combination of risk status and observed academic difficulties should immediately indicate the provision of a prereferral intervention to the target child, rather than waiting for further educational difficulties to manifest themselves. The provision and careful monitoring of such interventions has been shown to result in academic improvement (Quiroga, Lemos-Britton, Mostafapour, Abbott, & Berninger, 2002), which should minimize the extent and/or depth of the problems that could occur later.

Monitoring, rather than placement, would incur no additional costs, nor would it lead to the improper labeling of any child. Rather it would simply serve as an early warning signal of potential, not certain, academic problems that might eventually require remedial instruction or for some, a special education placement.

What factors increased the risk for achieving a score of less than 21? A relation between poor cognitive performance and minority status has been previously demonstrated in similar screening studies (e.g., Scott, Fletcher, Jean-Francois, et al., 1998; Scott et al., 2003) and minority status had a major negative im-

pact on performance again in this study. Hispanic and Black children were more than five times more likely to earn a score below the cut and five times more likely to be in special education compared to White children. It should be noted, however, that both SLD and EMH are low prevalence conditions so that overall, only 3.4% of the Black sample and 4% of the Hispanic sample actually ended up in special education.

A high percentage of those Hispanic and Black children with scores below the cut were also poor achievers. This outcome was more prevalent overall than was a placement in special education, with 18% of the total Hispanic sample and 36% of the total Black sample meeting the criterion for poor achievement in third grade.

Negative impact of minority status was mitigated for Hispanic children on some outcome measures if they were tested in English rather than Spanish or both languages. Children from the Hispanic/English group were less likely to have a kindergarten screening score below the cut and if they did have a score below the cut, were much less likely to be a poor achiever in third grade. Indeed, the Hispanic/English sample looked much more like the nonminority, White group, than did either of the other two minority groups.

Therefore, any Hispanic child who chooses (needs) to be tested in Spanish or both languages and who achieves a score below the cut should be rapidly responded to at the first sign of any achievement problems. A prereferral intervention geared to foster relevant cognitive processing skills and to improve English language performance and comprehension might be a cost effective approach to use with this particular group.

The increased risk for obtaining a screening score below the cut observed for the Hispanic/Spanish and Hispanic/Spanish-both groups was not related to educational status in third grade. Of the five Hispanic children in special education, none was tested in Spanish or both languages. All were tested in English. Absence of a negative impact of testing in other than English on this measure is probably related to the fact that the total number of children who were not tested in English was only 24 compared to the 101 who were tested in English. Given the low prevalence of a special education placement in the two categories examined, 4.5% over grades one through six, it is not surprising that all five Hispanic children who were in special education in third grade came from the larger sample.

The relation between SES and screening performance proved to be a major confound of the race/ethnicity effect. Nearly half the children in special education were Black and of those Black children with scores below the cut, 68% were poor achievers. However, it is likely that the poor performance of this group is related not to race per se, but to the fact that 77% of the Black sample lived in low SES catchment areas. Not one Black child in this sample was attending a high SES school. The confounding of minority status and low SES is consistent with other reported findings (Halle et al., 1997) as is the relation between low SES and poor achievement (Hill, 2001). All the characteristics that project to an expectation of poor academic performance are most likely to correlate with poverty, e.g., lower education in the parents, less verbal interactions between parents and children, lack of or little written material in the home. It would seem reasonable to conclude that the poorer performance of this particular minority group was drastically affected by factors associated with low SES.

Similarly, the superior performance of the Hispanic/English group compared to the Hispanic/Spanish or Hispanic/Spanish-both groups is partly confounded by SES. The percentage of the Hispanic/English group who lived in a high SES catchment area was twice as large as the percentage of the Hispanic/Spanish-both group. In contrast, the percentage of the Hispanic/Spanish-both group who were attending low SES schools was nearly three times as large as the percentage of children from the Hispanic/English sample who lived in a low SES catchment area.

Overall, it must be noted that on nearly all measures the Hispanic/Spanish and Hispanic/Spanish-both groups also performed more poorly than the Black group.

These data can be contrasted with a much poorer educational outcome for the Hispanic/Spanish and Hispanic/Spanish-both groups seen in the preschool predictive study (Scott & Delgado, 2003) when more than 84% of children in both groups were attending low SES schools. Clearly SES is a confounding vari-

able that must be taken into account when comparing racial/ethnic groups.

Since this screening set was selected based on the specific performance of this particular group of children, a cross validation study is necessary as are a number of other evaluative studies, such as an assessment of the test-retest reliability of the screening set. However, these data do show that it is possible to identify nearly all kindergarten children with only mild learning problems who will require special education services at a later date and a little more than a third of the children who will have achievement problems later in school.

Early identification of kindergarten children at risk for academic problems regardless of severity is critical to successful remediation efforts. This brief screening set, which is easy to administer and takes little time, will allow one to target a subset of children, the majority of whom are at greatest risk for a poor educational outcome. Monitoring the actual performance of this limited subset of children in order to detect any early signs of learning difficulties would be an inexpensive method of attacking a major educational problem.

References

Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1996). *Scales of Independent Behavior - Revised.* Chicago: The Riverside Publishing Company.

Dunst, C. J., Trivette, C. M., & Cross, A. H. (1986). Mediating influences of social import: Personal, family, and child outcomes. *American Journal of Mental Deficiency, 90,* 403–417.

Feinstein, A. R. (1976). On the sensitivity, specificity, and discrimination of diagnostic tests. *Pharmacology & Therapeutics, 17,* 104–116.

Funk, S. G., Sturner, R. A., & Green, J. A. (1986). Preschool prediction of early school performance: Relationship of McCarthy Scales of Children's Abilities prior to school entry to achievement in kindergarten, first, and second grades. *Journal of School Psychology, 24,* 181–194.

Gredler, G. R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools, 34,* 99–106.

Greenfield, D. B., & Scott, M. S. (1985). A cognitive approach to preschool screening. *Learning Disabilities Research, 1,* 42–49.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1996). Learning disability, low achievement, and mild mental retardation: More alike than different? *Journal of Learning Disabilities, 29,* 570–581.

Halle, T. G., Kurtz-Costes, B., & Mahoney, J. L. (1997). Family influences on school achievement in low income, African American children. *Journal of Educational Psychology, 89,* 527–537.

Hill, N. E. (2001). Parenting and academic socialization as they relate to school readiness: The roles of ethnicity and family income. *Journal of Educational Psychology, 93,* 686–697.

Holtzman, W. H. Jr., & Wilkinson, C. Y. (1991). Assessment of cognitive ability. In E. V. Hamayan & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 220–245). Austin: Pro-Ed.

Jacob, S., Snider, K. P., & Wilson, J. F. (1988). Validity of the Dial-R for identifying children with special educational needs and predicting early reading achievement. *Journal of Psychoeducational Assessment, 6,* 289–297.

Jean-Francois, B., Urbano, R. C., & Scott, M. S. (1999). *Test-retest reliability of the new cognitive preschool screening tests.* Unpublished doctoral dissertation, University of Miami.

Lazar, I., & Darlington, R. (1982). Lasting effects of early education: A report from the Consortium for Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (2–3, Series No. 195).

Lenkarski, S., Singer, M., Peters, M., & McIntosh, D. (2001). Utility of the Early Screening Profiles in identifying preschoolers at risk for cognitive delays. *Psychology in the Schools, 38,* 17–24.

Lichtenstein, R. (1981). Comparative validity of two preschool screening tests: Correlational and classificational approaches. *Journal of Learning Disabilities, 14,* 68–72.

Lichtenstein, R., & Ireton, H. (1984). *Preschool screening.* Orlando, FL: Grune & Stratton.

Lichtenstein, R., & Ireton, H. (1991). Preschool screening for developmental and educational problems. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (2nd ed., pp. 486–513). Boston: Allyn & Bacon.

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities, 34,* 136–151.

Mantzicopoulos, P. Y., & Morrison, D. (1994). Early prediction of reading achievement. *Remedial and Special Education, 15,* 244–251.

McIntosh, D. E. (1999). Identifying at-risk preschoolers: The discriminant validity of the Differential Abilities Scales. *Psychology in the Schools, 36,* 1–10.

Meisels, S. J. (1989). Can developmental screening tests identify children who are developmentally at risk? *Pediatrics, 83,* 578–585.

Mercer, C. D., Algozzine, B., & Trifiletti, J. (1988).

Early identification: An analysis of the research. *Learning Disability Quarterly, 11,* 176–188.

Public Law 99–457.(1986). *Education of the Handicapped Act Amendments of 1986.* Washington, DC: Department of Health, Education and Welfare.

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness and beginning reading in Spanish speaking ESL first graders. *Journal of School Psychology, 40,* 85–111.

Rafoth, M. A. (1997). Guidelines for developing screening programs. *Psychology in the Schools, 34,* 129–142.

Salvia, J., & Ysseldyke, J. E. (1991). *Assessment* (5th ed.). Boston: Houghton Mifflin. Scott, K. G., Urbano, J. C., & Boussey, C. (1991). Long term psychoeducational outcome of prenatal substance exposure. *Seminars in Perinatology, 15,* 317–323.

Scott, M. S., & Delgado, C. F. (2003). Using educational placement in third grade to select and validate a preschool screening measure. *Psychology in the Schools, 40,* 565–582.

Scott, M. S., Deuel, L. S., Claussen, A. H., & Sanchez, M. (1993). Identifying young children with mild cognitive deficiencies. *Diagnostique, 19,* 335–359.

Scott, M. S., Deuel, L. S., Urbano, R. C., Fletcher, K. L., & Torres, C. (1998). Evaluating the initial version of a new cognitive screening test. *Education and Training in Mental Retardation and Developmental Disabilities, 33,* 280–289.

Scott, M. S., Fletcher, K. L., & Deuel, L. S. (1998). The effects of intelligence on the identification of young children with learning disabilities. *Learning Disabilities Research & Practice, 13,* 81–88.

Scott, M. S., Fletcher, K. L., Jean-Francois, B., Urbano, R. C., & Sanchez, M. (1998). New screening tests to identify young children at risk for mild learning problems. *Journal of Psychoeducational Assessment, 16,* 302–314.

Scott, M. S., Greenfield, D. B., & Sterental, E. (1986). Abstract categorization as a predictor of learning disability classification. *Intelligence, 10,* 377–387.

Scott, M. S., Perou, R., Greenfield, D. B., & Swanson, L. (1993). Rhyming skills: Differentiating among the mildly mentally retarded, learning disabled and normally achieving students. *Learning Disabilities Research & Practice, 8,* 215–222.

Scott, M. S., Tu, S., & Fletcher, K. L. (2003). Cross validating a new preschool screening test. *Education and Training in Developmental Disabilities, 38,* 167–178.

Shonkoff, J. P., & Hauser-Cram, P. (1987). Early intervention for disabled infants and their families: A quantitative analysis. *Pediatrics, 80,* 650–658.

Simmons, D. C., & Kameenui, E. J. (1990). The effect of task alternatives on vocabulary knowledge: A comparison of students with and without learning disabilities. *Journal of Learning Disabilities, 23,* 291–297.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Stebbins, M. S., & McIntosh, D. E. (1996). Decision-making utility of the Bracken Basic Concept Scale in identifying at-risk preschoolers. *School Psychology International, 17,* 293–303.

Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities, 27,* 34–50.

Tramontana, M. G., Hooper, S. R., & Selzer, S. C. (1988). Research on the preschool prediction of later academic achievement: A review. *Developmental Review, 8,* 89–146.

van Kraayenoord, C. E. (1983). Preschool screening for learning disabilities: Is it warranted? *New Zealand Journal of Educational Studies, 18,* 48–58.

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques. *Public Health Reports, 62,* 1432–1449.

Received: 28 April 2004
Initial Acceptance: 29 June 2004
Final Acceptance: 20 December 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 390–400

Effectiveness of Constant Time Delay on Teaching Snack and Drink Preparation Skills to Children with Mental Retardation

Funda Bozkurt and Oguz Gursel
Anadolu University

Abstract: A multiple probe design with probe conditions across behaviors was used to evaluate effectiveness of constant time delay on teaching snack and drink preparation skills to children with mental retardation. In addition, generalization effects across settings, time, and materials, and maintenance effects were examined. Three students between the ages of 14 and 17 enrolled in a special education classroom served as subjects. Distributed trial format was used in the probe, maintenance, and generalization sessions to assess target behaviors. Results indicated that constant time delay was effective in teaching subjects to make a sandwich, to prepare a hot drink, and to serve these foods and drinks.

Currently, educational researchers are developing effective teaching methods based on information from experimental studies. Research based methods contribute to the quality of teaching by providing researchers and educators with insights into methods that are effective in teaching individuals displaying different learning abilities (Tekin & Kırcaali-Iftar, 2001).

There is a substantial body of research in relation to effective teaching. One method is errorless teaching (Schuster & Griffen, 1990; Schuster et al., 1998; Tekin & Kircaali-Iftar, 2001). According to this teaching approach, acquisition of skill and concept is sourced from the positive responses and practices in which learners engage in, not from the errors committed (Tekin, 1999; Wolery, Bailey, & Sugai, 1998).

Tekin and Kircaali-Iftar (2001) state that there are various instructional procedures in errorless teaching methods, one of them is known as constant time delay (CTD). CTD is considered to be one of the effective procedures that could be be used to teach individuals with mental retardation (Schuster et al., 1998; Wolery, Ault, & Doyle, 1992).

CTD procedure is conducted in two phases: (a) 0 s time delay trials, and (b) constant time delay trials (Ault, Gast, Wolery, & Doyle, 1992; Schuster & Griffen, 1990; Schuster et al., 1998; Tekin & Kircaali-Iftar, 2001).

In 0 s time delay trials, target stimulus and prompt are provided simultaneously. Following the provision of target stimulus, prompts are provided in order for the learner to respond correctly (Ault et al., 1992; Schuster & Griffen, 1990; Tekin & Kircaali-Iftar, 2001).

Constant time delay trials, on the other hand, is a phase in which the prompt is faded by inserting a fixed time such as 4 or 5 s between the target stimulus and prompt. For the following instruction sessions, this waiting time is inserted and these sessions are called constant time delay sessions (Tekin & Kircaali-Iftar, 2001). There is a great deal of research demonstrating effectiveness of CTD procedures on teaching both discrete and chained skills to individuals with various disabilities

The study was completed by the first author in partial fulfillment for the requirements of a Master's of Art degree in Special Education in Anadolu University, Eskisehir, Turkey. Authors are grateful to Dr. Elif Tekin-Iftar, Research Institute For The Handicapped, for her insightful review and feedback, and to Serife Yucesoy, Anadolu University, for collecting reliability data. This study was presented at the International Conference on Speciel Education in Antalya, Turkey in July 2001 and at the Association for Behavior Analysis (ABA) Annual Convention in Toronto, Canada in May 2002. Correspondence concerning this article should be addressed to Oguz Gursel, Egitim Fakultesi, Ozel Egitim Bolumu, Anadolu Universitesi, Eskisehir, TURKEY, 26470. Email: ogursel@anadolu.edu.tr

and from different age groups (Schuster et al., 1998; Tekin & Kircaali-Iftar, 2001).

Miller and Test (1989) compared the effectiveness and effiency of CTD and most to least prompting on teaching laundry skills. They reported that both procedures were equally effective, however, CTD was more efficient on training time to criterion and number of incorrect responses to criterion.

In another study conducted by Miller and Test (1989), the effectiveness and efficiency of decreasing prompt and CTD were compered. Both procedures were found to be equally effective, however CTD was found to be more efficient than decreasing prompt in terms of training time and student error rate (Schuster & Griffen, 1991; Schuster et al., 1998).

Researchers frequently rely on CTD when teaching chained skills. Among these are cooking skills (Hall, Schuster, Wolery, Gast, & Doyle, 1992; Schuster, Gast, Wolery, & Guiltinan, 1988), banking skills (Donnell & Ferguson, 1989), washing clothes (Miller & Test, 1989), leisure skills (Tekin-Iftar et al., 2001), housework skills (Wolery, Ault, Gast, Doyle, & Griffen, 1990, 1991), first aid skills (Gast, Winterling, Wolery, & Farmer, 1992) and personal skills (Ault et al., 1992; Schoen & Sivil, 1989; Schuster & Griffen, 1991; Schuster et al., 1998).

Schuster et al. (1998) reviewed 20 research reports utilizing CTD for teaching chained skills. In 13 studies, the focus was on effectiveness of CTD and in two studies upon delivering instructive feedback in which CTD is conducted. In the remaining five studies, most to least prompting and least to most prompting are compared with CTD. Of 81 subjects who participated in these studies, 65 met the criterion in skill teaching. With respect to measures of effectiveness, 19 studies reported efficiency measures, such variables as number and percente of errors, number of sessions, exact time of teaching and error types have also been examined. In studies, CTD was found to be effective in the teaching of chained skills to individuals with different disabilities between the ages of 2 and 48 (Schuster et al., 1998).

In comparison studies (McDonnell, 1987; Schoen & Sivil, 1989; Schuster & Griffen, 1991; Schuster et al., 1998; Wolery et al., 1990) it is reported that least to most prompting and CTD were effective, however, CTD was more efficient than least to most prompting in terms of training time, number of trials and sessions toward criterion and number and percent of incorrect responses.

Furthermore, Tekin and Kircaali-Iftar (2001) state that CTD provides learners with more opportunities for reinforcement. They also point out that this method requires less preparation on the part of the teacher and that, for this reason, it is considered to be more efficient.

As is clear that CTD is effective in teaching both discrete and chained-skills to children with mental retardation. In Turkey, it appears that teachers experience difficulty in helping children with mental retardation to acquire both discrete skills and chained-skills. Tekin and Kircaali-Iftar (2001) compared the effectiveness and efficiency of CTD and simultaneous prompting on teaching receptively identifying animal names to children with mental retardation. In their study, the focus was upon teaching leisure-time activities. However, there is a lack of research investigating effectiveness of errorless teaching methods and CTD in the teaching of kitchen skills in Turkey.

In order to be able to lead an independent life, it is important for children with mental retardation to acquire kitchen skills. Children need to learn how to cook, how to preserve food and drink, how to clean the kitchen, the nutrition value of foods, and so on. It is important to teach independent living skills, and particularly cooking skills, to individuals with mental retardation.

The purpose of this study is to examine the effectiveness of CTD on teaching of the skills 'making a sandwich, preparing a hot drink and serving' to three students with mental retardation. The research questions are: (a) Is CTD effective on the teaching of food and preparation and serving skills to three students with moderate mental retardation? (b) Are the students be able to maintaine the acquired kitchen skills 2 and 4 weeks after the training? (c) Are the students be able to generalize the acquired kitchen skills across different contexts, time and materials?

Method

Participants and Setting

Three students with moderate mental retardation (2 female, 1 male) 14, 16 and 17 years old were particicpants. All participants attended a literacy class at a rehabilitation center for two-years. Students in these classes take literacy, math, science, physical education, technical education, music, and art courses. None of the participants had a history with CTD.

Prerequisite skills for the participants were: (a) ability to recognize objects, (b) ability to follow verbal instructions, (c) ability to imitate prompts, (d) ability to use their two hands in cooperation, (e) ability to stand still at least 5 minutes during instruction, and (f) ability to wait for the prompts. In order to understand whether participants had the given prerequisite skills, the class teacher was interviewed, and the researchers conducted unsystematic observations before the study.

Demet (14 years old) and Umut (17 years old) were diagnosed with moderate mental retardation. Tugce was a 16 year-old student with Down syndrome. These three students were able to behave independently as to self-care skills, gross motor and fine motor skills, psychomotor skills, and language skills. They were also be able to understand and perform instructions consisting of three words and were able to read and write half-page texts. They were able to recognize such objects, food and drinks as dish, spoon, fork, plate, glass, pot, kettle, cheese, sugar, water and oralet. They were able to pour water into a jug and glass and to use a spoon and fork. Demet was given such responsibilities as tidying her room and cleaning in contrast to Tugce and Umut.

In the present study, the probe and instruction sessions were conducted between 11:00 and 12:00 on Monday, Wednesday and Friday. One session was conducted within a day. Maintenance sessions were carried out 2 and 4 week after training. These sessions were held between 11:00 and 12:00 at lunch time. With regard to the generalization sessions conducted immediately after the maintenance sessions, they were conducted between 2:00 and 3:00 p.m. All sessions were conducted in an individual teaching format.

Researchers

The first author of the study, a research assistant attending an MA programme, had five years experience teaching students with mental retardation. The second author of the study was an assistant professor in the field of special education. Reliability data was collected by two independent observers, the second author and a graduate student in department of special education.

Materials

In instruction sessions, two sets of materials were used, one for the student and one for the researcher. In probe sessions, only one set of materials was used for the student. Materials used in probe and instruction sessions were bread, cheese, sliced tomato, plastic plate and fork for the skill of making a sandwich; sugar, oralet, water, kettle, tray, plastic glass and spoon for the skill of preparing a hot drink, and a sandwich in a dish, a glass of hot oralet, tray and napkin for the skill of serving. In generalization sessions, students were given baguette, lettuce, sausage, glass dish and metal fork for the skill of making a sandwich; plate, sugar, sour-cherry oralet, water, kettle, glass and metal spoon for the skill of preparing a hot drink and a previously-prepared sandwich in a dish, a glass of hot oralet, a colorful tray and napkin for the skill of serving.

Analysis of the Skills

Task analyses for the target behaviors, making a sandwich, preparing a hot drink and serving, were developed by the authors as a forward chaining procedure, and the skills were taught simultaneously. After skills were analyzed, each step was recorded, experts in Special Education were consulted, and essential corrections were made. The skill of making a sandwich and serving consist of 11 steps and the skill of preparing a hot drink was 14 steps (see Table 1).

Experimental Design

A multiple probe design with probe conditions across behaviors was replicated across

TABLE 1

Task Analysis of Making a Sandwich, Preparing Hot Drink and Serving

Making a Sandwich	*Preparing a Hot Drink*	*Serving*
1. take get the dish.	1. plug the socket of the kettle in.	1. get the tray.
2. take get the bread.	2. open the lid of the kettle.	2. put the tray on the counter.
3. put the bread in the dish.	3. pour a glass of water into the kettle.	3. get the dish with the sandwich.
4. separate the bread in half.	4. shut down the lid of the kettle.	4. put the sandwich dish on the tray.
5. put the upper part of the bread in the dish.	5. push the button which reads 1 to start the kettle.	5. get the glass.
6. take get the cheese.	6. open the lid of the oralet jar.	6. put the glass on the tray.
7. put the cheese on the bottom part of the bread.	7. put two or three tea spoons of oralet in the glass.	7. get napkins.
8. take get the fork.	8. open the lid of the sugar jar.	8. puts the napkins on the tray.
9. get the tomato.	9. put two or three tea spoons of sugar in the glass.	9. get the tray on the counter.
10. put the tomato in the bread.	10. take the kettle when it is off.	10. get the tray to the table.
11. place the upper part of the bread on the bottom part of it.	11. pour the water in the kettle into a glass.	11. put the tray on the table.
	12. leave the kettle where s/he taken it.	
	13. unplug the socket of the kettle.	
	14. stir the ingredients in the glass with a spoon.	

subjects. Experimental control was built in when the subject was responding at or near baseline levels during full probe conditions before the intervention was introduced and criterion was reach of only after the intervention was introduced for each subject (Tekin & Kircaali-Iftar, 2001).

General Procedures

Probe, instruction, maintenance, and generalization sessions were held for each learner. Instruction sessions were an individual teaching format, and the skills of making a sandwich, preparing a hot drink and serving were taught simultaneously. The following task instructions were given to students: (a) For making a sandwich, 'Please, make a sandwich.' (b) For preparing hot drink, 'Please, prepare a hot drink.' (c) For serving, 'Please, serve what you have prepared.'

Also, regardless of whether students correctly responded to the prompts, they were told that they could eat half of what they had prepared and that they could have all that they had prepared if they could successfully complete all skill steps alone.

Probe Sessions

Data was gathered by means of multiple opportunity method in a distributed trial format. In a multiple opportunity method, when the student incorrectly responded to the prompt, the researcher secretly completed the given step instead of the student. Data was collected for each basic skill. Students were not provided with any prompt to complete the skill steps, however, they were reinforced through continuous (CRF), variable ratio (VR3) and fixed ratio (FR12, FR14) schedules. Students' incorrect responses were ignored. If the learner failed to respond to the prompts, this was considered an incorrect response. Students were observed as to whether or not they correctly completed each step after the target prompts.

Probe sessions were held as follows: To draw the student's attention to the skill of making a sandwich, the researcher asked, 'Now, we will make a sandwich together. Are you ready?'

When the student said they were ready, the researcher verbally praised the student, 'Great! You are ready'. The student was then given the skill instruction, 'Please make a sandwich'. Following this instruction, the student was given 4 s, to complete the first step of the skill, s/he and the response was recorded. The student was then observed for completion of successive steps within 4 s. Once the student successfully completed all steps of the skill verbal reinforcement, 'Thank you for your participation' was given.

Baseline data were gathered until stable for each the student. Subsequent to this procedure, probe sessions were held to collect probe data. At least three probe sessions were conducted until stable data were obtained. This procedure was conducted for each skill and each student.

Intervention Sessions

Following above procedures, sessions with 0 s time delay trials were conducted until students were at 100% correct responses to the given prompts. Then, 4 s time delay trials were conducted until students were at 100% correct responses to the given prompts in at least three consequtive sessions.

For making a sandwich and serving skills, only one session was held with each student. For the skill of preparing a hot drink, one session was held with Demet and Umut and two sessions with Tugce. All sessions were 0 s time delay. As in probe sessions, the researcher asked, 'Now, we will make a sandwich together. Are you ready?' Once students said they were ready, they were told 'Please make a sandwich'. The student was then provided with the correct response and asked to do the same. Correct responses resulted in verbal reinforcement. When students responded incorrectly they were interrupted and corrected. The researcher then completed the given step using verbal and model prompts. If the student was unable to respond to the prompt, this was considered an incorrect response. The researcher completed the given step using verbal and model prompts. Once the students successfully completed all steps of the skill, they were verbally reinforced.

The same procedures were followed in the 4 s time delay sessions. The student's attention was secured and given the skill instruction. The researcher waited 4 s for the student to complete the first step. If the student correctly completed the step within 4 s (unprompted correct response), verbal praise was given. If the student incorrecly responded (unprompted incorrect response), the researcher helped the student respond correctly with simultaneous verbal and model prompts. If the student was unable to respond within 4 s, the researcher provided simultaneous verbal and model prompts. If the student than correctly responded (prompted correct response), verbal praised was given. If the student incorrectly responded despite the given prompts (prompted incorrect response), the researcher completed the given step and explained what he was doing. If the student did not respond (no response), this was considered an incorrect response. The researcher completed the step for the student and moved on to the following step. Other steps of the skills were completed in the same way. When students completed all steps of the skill, they were verbally reinforced, 'Thank you for your participation'. In addititon, they were allowed to eat what they had prepared (FR12, FR14). The same procedures were followed for each different skill and each student.

Students' unprompted and prompted correct responses were reinforced with CRF, VR3, FR12, and FR14 schedules. When students incorrectly reached the unprompted response, they were told to wait for the prompt, and the instruction was repeated. When they incorrectly reacted to the prompted response, they were interrupted and corrected. Students' no responses were ignored, and they were given the prompted correct response.

Maintenance and Generalization Sessions

After 2 and 4 weeks effects of instruction were examined. Maintenance and generalization sessions were conducted the same way as probe sessions. Steps completed successfully, were reinforced on FR11 and FR14 schedules. Participation was reinforced on a VR3 schedule. Reinforcements were faded in maintenance sessions. In the first session, the students were given food and verbally praised. In

the second session, they were verbally praised.

Generalization sessions were held immediately after maintenance sessions. Sessions were the same as probe sessions except different context and different materials were used. Sessions were conducted at the personnel kitchen of the special education center between 2:00 and 3:00 p.m. Materials used were glass dishes, bread, sausages, lettuce, fork, glass, plate, tea spoon, sugar, oralet, water, kettle, and napkin. Students were continously reinforced via verbal and tangible reinforcement.

Reliability Analysis

Two observers watched all videorecorded sessions and collected data as to both procedural and interobsever reliability.

Reliability data was gathered at least across 20% of all sessions. Interobsever reliability was 100% for skills of making sandwich and preparing a hot drink in probe sessions and between 97%-100% for the skill of serving. In instruction sessions, reliability was between 91%-100% for the skills of making sandwich and serving and 100% for the skill of preparing a hot drink. In maintenance and generalization sessions 100% agreement was found for the skills of making sandwich and preparing a hot drink, and 91%-100% for the skill of serving.

In order to ensure procedural reliability one observer checked whether the researcher: (a) had materials ready, (b) secured attention, (c) gave task direction, (d) provided prompted corrected reponse, (e) gave correct response, and (f) waited 4 s. Procedural reliability was 100%.

Results

Effectiveness Data

Results indicate that CTD was effective in teaching students how to make a sandwich, how to prepare a hot drink, and how to serve what they had prepared. Figures 1, 2, and 3, present the results of the instruction for each participant on each of the skills in each session.

Maintanence and Generalization

Maintenance sessions infer a more permanent effect of the instruction on students' acquired skills. Demet and Umut achieved 100% retention for each skill, while Tugce achieved 100% retention only for preparing a hot drink and serving in two maintenance sessions. With regard making a sandwich, she achieved 100% retention in the first maintenance session and 91% in the second.

In generalization sessions, each of the students was able to generalize skills across different contexts and materials. For making a sandwich, Demet and Tugce achieved 73% generalization and Umut 100%. For preparing a hot drink, Demet and Umut were 100% and Tugce 93%. For serving, all students were at 100% generalization.

Discussion

The present study examined effectiveness of CTD on teaching snack and drink preparation, and serving skills to children with mental retardation.

Findings indicated that CTD was effective in teaching these skills. In probe sessions, each of the students successfully achieved the steps in each skill. Similarly, in maintenance sessions, permanent effects of the instruction were observed. Further, all subjects were able to generalize skills across different contexts with different materials. These findings concur with those in other similar studies. Harden and Zane (as cited in Schuster et al., 1998), Tekin-Iftar et al. (2001) and Wolery et al. (1992); for example, found CTD to be effective in the teaching of various skills to children with mental retardation. Schuster et al., in their review of 20 research articles, concluded that all studies demonstrated effectiveness of CTD.

In order to teach the skills, 0 s and 4 s time delay trials were conducted. These two procedures were included in the study due to the chained-skill instruction. Although participants correctly responded 100% to the skill direction in the 0 s time delay sessions, they could not display similar performance in sessions with 4 s time delay. In these sessions, they could not give correct responses to the target stimulus, but only to prompted correct

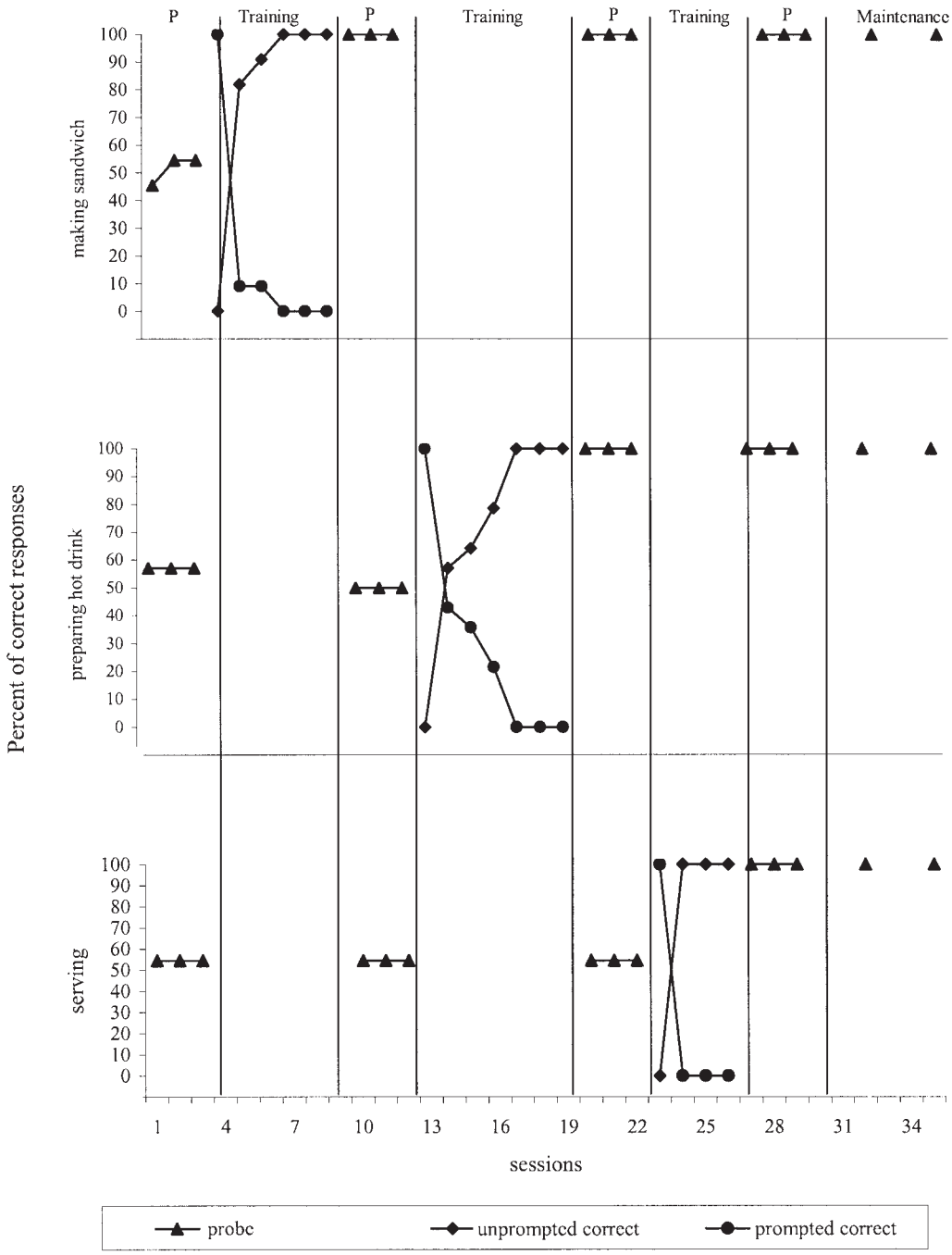

Figure 1. Percent of correct responses for Demet across chained skills taught with constant time delay.

responses. What this implied was a higher number of sessions and longer instruction time. Participants' correct responses before and after the prompts were reinforced in the same way, and responses given before the prompt were considered to be correct. It

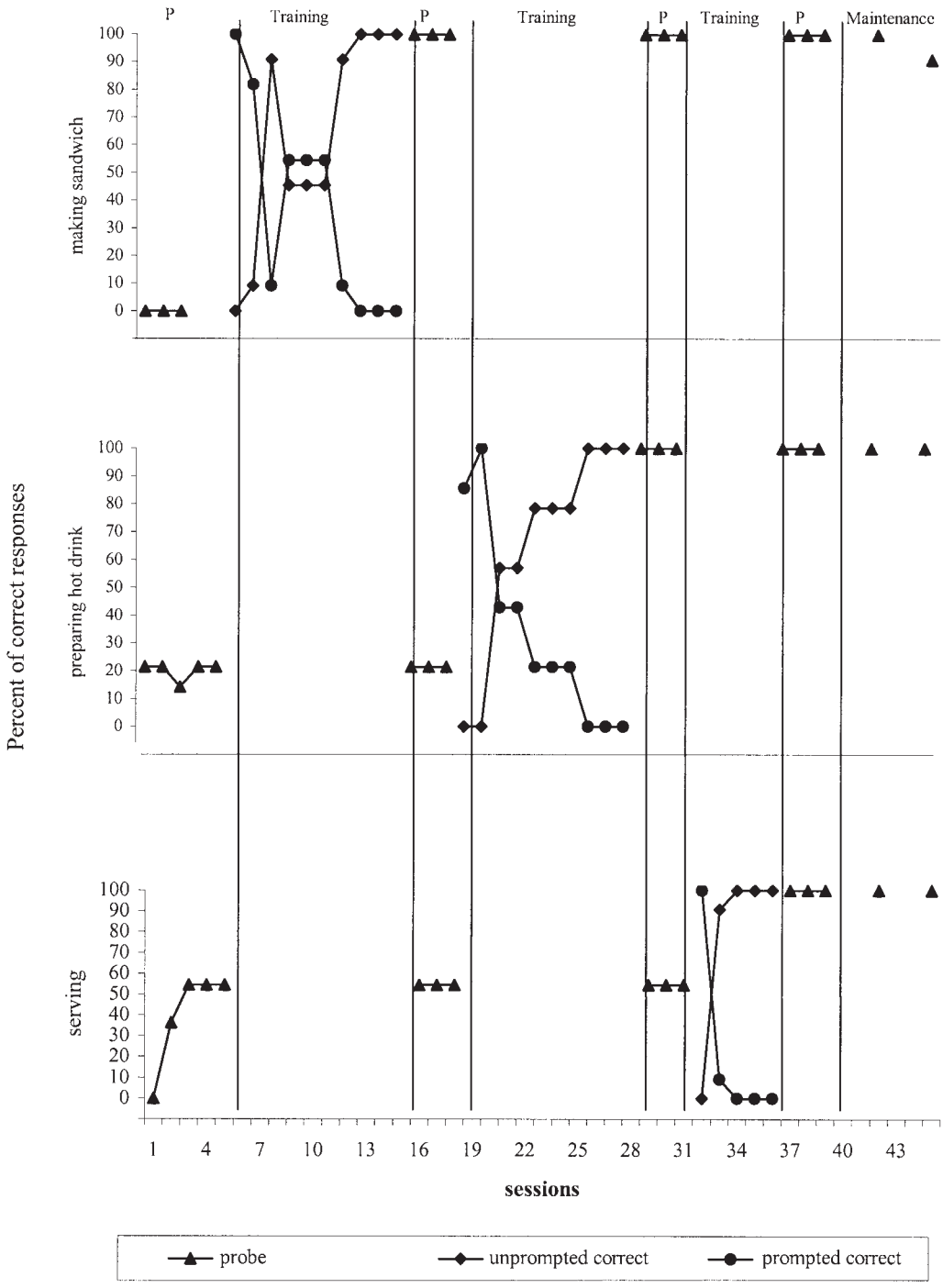

Figure 2. Percent of correct responses for Tugce across chained skills taught with constant time delay.

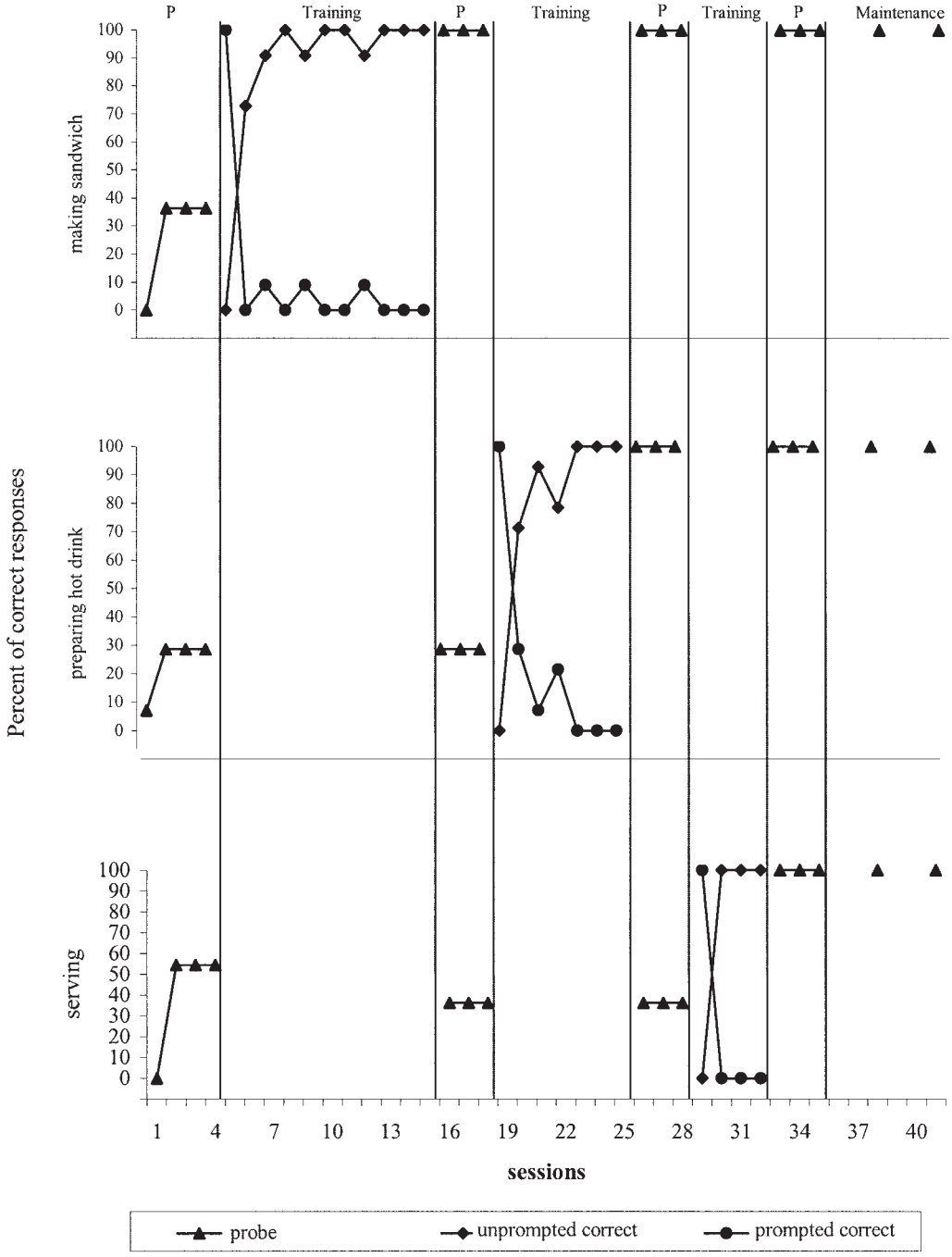

Figure 3. Percent of correct responses for Umut across chained skills taught with constant time delay.

could be that reinforcing both types of correct responses in the same way lengthened instruction time and that differential reinforcement would be appropriate for correct responses preceding prompts in order to meet the criterion within a shorther period of time. Along

with use of reinforcements for correct responses, the incorrect responses were corrected, and arguably, error correction prevented participants from giving incorrect responses. Also, none of the participants were given wait training instruction. This might be one of the reasons for a short period of instruction.

Future studies could undertake research into the effectiveness of CTD in different settings with a focus upon teaching different chained skills, with use of different prompts, and with instruction sessions in which different 0 s and 4 s time delay procedures are employed. Also, future studies might tackle the question of whether forward chaining, backward chaining and the total skill format are effective. Another segment of future studies could focus upon effectiveness of instruction sessions in which participants are given ample time to complete each step of the skills and compare CTD with another errorless instruction procedure. Further studies could also deal with the use of CTD for the teaching of one-step skills and error types emerging both in discrete skills and chained skills. In conclusion, in view of this study and previous studies, it could be argued that CTD is effective in teaching chained skills to individuals with mental retardation and is a procedure that is easily applied.

References

Ault, M. J., Gast, D. L., & Wolery, M. (1988). Comparison of progressive and constant time delay procedures in teaching community-sign word reading. *American Journal on Mental Retardation, 93,* 44–53.

Ault, M. J., Gast, D. L., Wolery, M., & Doyle, P. M. (1992). Data collection and graphing method for teaching chained tasks with the constant time delay prodecure. *Teaching Exceptional Children, 24,* 28–33.

Gast, D. L., Winterling, V., Wolery, M., & Farmer, L. A. (1992). Teaching first-aid skills to students with moderate handicaps in small group instruction. *Education and Treatment of Children, 15,* 101–124.

Hall, M. G., Schuster, J. W., Wolery, M., Gast D. L., & Doyle, P. M. (1992). Teaching chained skills in a non-school setting using a divided half instructional format. *Journal of Behavioral Education, 2,* 257–274.

McDonnell J. (1987). The effects of time delay and increasing prompt hierarchy strategies on the acquisition of purchasing skills by students with severe handicaps. *The Association for Persons with Severe Handicaps, 12,* 227–236.

McDonnell, J., & Ferguson, B. A. (1989). Comparison of time delay and decreasing prompt hierarchy strategies in teaching banking skills to students with moderate handicaps. *Journal of Applied Behavior Analysis, 22,* 85–92.

Miller, U. C., & Test, U. W. (1989). A comparison of constant time delay and most-to-least prompting in teaching laundry skills to students with moderate retardation. *Education and Training in Mental Retardation, 24,* 363–370.

Schoen, S. F., & Sivil, E. O. (1989). A comparison of prodecures in teaching self-help skills: Increasing asistance time delay, and observational learning. *Journal of Austism and Developmental Disorders, 19,* 57–72.

Schuster, J. W., Gast, D. L., Wolery, M., & Guiltinan, S. (1988). The effectiveness of a constant time delay prodecure to teach chained responses to adolescents with mental retardation. *Journal of Applied Behavior Analysis, 21,* 169–178.

Schuster, J. W., & Griffen, A. K. (1990). Using time delay with task analyses. *Teaching Exceptional Children, 22,* 49–54.

Schuster, J. W., & Griffen, A. K. (1991). Using constant time delay to teach recipe following skills. *Education and Training in Mental Retardation, 26,* 49–53.

Schuster, J. W., Morse, T. E., Ault, M. J., Doyle, P. M., Crawford, M. R., & Wolery M. (1998). Constant time delay with chained tasks: A review of the literature. *Education & Treatment of Children, 21,* 33–75.

Tekin, E. (1999). Yanlıþssiz ogretim yontemleri (Errorless teaching procedures in special education). *Ozel Egitim Dergisi (The Journal of Special Education), 2,* 87–102.

Tekin, E., & Kircaali-Iftar, G. (2001). *Ozel egitimde yanlıssız ogretim yontemleri (Errorless teaching procedures in special education).* Ankara, Turkey: Nobel Basin Yayin.

Tekin-Iftar, E., Kircaali-Iftar, G., Birkan, B., Uysal, A., Yıldırım, S., & Kurt, O. (2001). Using constant time delay to teach leisure skills to children with developmental disabilities. *Mexican Journal of Behavior Analysis, 27,* 337–362.

Wolery, M., Ault, M. J., & Doyle, P. M. (1992). *Teaching students with moderate to severe disabilities: Use of response prompting strategies.* New York: Longman.

Wolery, M., Ault, M. J., Gast, D. L., Doyle, P. M., & Griffen, A. K. (1990). Comparison of constant time delay and system of least prompts in teach-

ing chained tasks. *Education and Training in Mental Retardation, 25,* 243–257.

Wolery, M., Ault, M. J., Gast, D., Doyle, P. M., & Griffen, A. K. (1991). Teaching chained tasks in dyads: Acquisition of target and observational behaviors. *The Journal of Special Education, 25,* 198–220.

Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching principles and procudures of applied behavior analiysis with exceptional students.* Boston: Allyn and Bacon.

Wolery, M., Doyle, P. M., Ault, M. J., Gast, D., Meyer, S., & Stinson, D. (1991). Effects of presenting incidental information in consequent events on future learning. *Journal of Behavioral Education, 1,* 79–104.

Received: 7 April 2004
Initial Acceptance: 1 June 2004
Final Acceptance: 25 October 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 401–410

Teaching Naming Relatives to Individuals with Autism Using Simultaneous Prompting

Nurgul Akmanoglu-Uludag and Sema Batu
Anadolu University

Abstract: This study examined the effectiveness of simultaneous prompting in teaching naming relatives to individuals with autism. Two 5.5 year old male participants who were diagnosed with autism were taught eight different relative names using simultaneous prompting. Maintenance and generalization data across materials, settings, and trainers were collected. Results revealed that simultaneous prompting was an effective way of teaching relative names to children with autism, and also that participants maintained and generalized the skills taught after training sessions were completed.

Imitation skills, initiating and continuing communication, functional communication skills, play skills, and role playing skills are primary skill areas to be taught to children with autism (Coyne, Nyberg, & Vandenburg, 1999; Hodgdon, 1999; Smith, 2001).

Functional communication skills are very essential parts of independent daily living. Children with or without disabilities need to learn functional communication skills in order to be a real part of the community independently. Children with autism especially need to be taught functional communication skills since they usually have very limited social interaction skills, communication, and play skills (Frost & Bondy, 2002; Kircaali-Iftar, 2003).

Knowing and using relative names is one of the ways of initiating communication within the family. In Turkey every relative has a different name to be used to call him or her. For example 'dayi' is uncle which is mother's brother, 'amca' is also uncle but it is used for father's brother; 'teyze' is aunt which is mother's sister, 'hala' is also aunt but it is used for father's sister, etc. Since everybody uses these names in order to talk to their relatives, one of the first things children learn are the relative names in his/her family. Teaching relative names is also part of the national preschool program in Turkey. Therefore, it is important for children with autism to learn to use relative names in order to initiate or continue communication with his/her relatives in daily lives and also to be accepted in the community.

Correspondence concerning this article should be addressed to Nurgul Akmanoglu-Uludag, Anadolu Universitesi, Engelliler Arastirma Enstitusu, Eskisehir, TURKEY 26470. Email: nakmanoglu@anadolu.edu.tr

Near-errorless teaching procedures are one of the ways of teaching communication skills to children with various disabilities. Simultaneous prompting is one of the most effective among near-errorless teaching procedures and also preferred as it is easily implemented by teachers and other practitioners (Tekin, & Kircaali-Iftar, 2001; Wolery, Ault, & Doyle, 1992).

During implementation of simultaneous prompting, the practitioner presents a controlling prompt (i.e., a prompt that ensures a correct response) simultaneously with the stimulus being taught. For example, the teacher shows a picture of a cat and says, "Tell me, what is on the picture?" and immediately responds "A cat." While implementing the procedure, the participant is not allowed to give an independent response because the practitioner presents the controlling prompt with the task direction simultaneously. Therefore, probe sessions play a very important role for determining the transfer of stimulus control for the participant (Parrott, Schuster, Collins, & Gassaway, 2000).

Simultaneous prompting is suggested for teachers and other practitioners because of

many advantages. One advantage is that the student does not need to be taught a waiting response. This is a skill that individuals with autism usually do not have (Koegel & Koegel, 1999). Another advantage is the similarity of sessions to each other during the teaching procedure. The practitioner does not need to change the time for responding, because all trials are conducted with a 0 s waiting interval between task direction and controlling prompt. Another advantage is that there is only one type of correct response for the learner; therefore, there is no need to differentiate the reinforcement procedure as in other errorless teaching strategies (Parrott et al., 2000; Tekin & Kircaali-Iftar, 2001).

Many studies examined effectiveness of simultaneous prompting in teaching both discrete skills ('pointing to the numerals' Akmanoglu & Batu, 2004; 'identifying occupations' Dogan & Tekin-Iftar, 2002; 'word recognition' Gibson & Schuster, 1992; 'word identification' Griffen, Schuster, & Morse, 1998; 'teaching science vocabulary words' Johnson, Schuster, & Bell, 1996; 'object identification' MacFarland-Smith, Schuster, & Stevens, 1993; 'receptive manual sign identification' Palmer, Collins, & Schuster, 1999; 'teaching sight words' Schuster, Griffen, & Wolery, 1992; 'teaching community signs' Singleton, Schuster, & Ault, 1995; 'reading grocery sight words' Singleton, Schuster, Morse, & Collins, 1999; 'teaching community signs' Tekin-Iftar, 2003; 'rebus symbols' Wolery, Holcombe, Werts, & Cipolloni, 1993) and chained skills ('teaching vocational task' Fetko, Schuster, Harley, & Collins, 1999; 'teaching hand washing' Parrott et al., 2000; 'academic skills' Riesen, McDonnell, Johnson, Polychronis, & Jameson, 2003; 'making juice' Schuster & Griffen, 1993; 'teaching dressing' Sewell, Collins, Hemmeter, & Schuster, 1998) to individuals with disabilities. Many of these studies were conducted with individuals with mental retardation (e.g., Dogan & Tekin-Iftar; Fetko et al.; Griffen et al.; Parrott et al.; Schuster & Griffen; Singleton et al., 1995; Sewell et al.; Singleton et al., 1999; Tekin-Iftar; Wolery et al., 1993, etc.). Only Akmanoglu and Batu examined the effectiveness of simultaneous prompting in teaching pointing to the numerals with individuals with autism. Therefore, there is a need to conduct studies examining effectiveness of simultaneous prompting on teaching various skills to individuals with autism.

The purpose of this study was to examine the effectiveness of simultaneous prompting on teaching naming relatives to children with autism. Research questions were: (a) Is simultaneous prompting an effective method for teaching naming relatives to children with autism? (b) Are participants going to maintain skills taught after one, two and four weeks? (c) Are participants going to generalize skills taught across materials, settings, and trainers? and (d) What do parents of participants think about the social validity of the study conducted with their children?

Method

Participants

Faruk and Orhan, two male students with autism, were the participants. Both were 5.5 years old and enrolled in the first author's classroom for children with autism at a university unit for children with developmental disabilities for the last five months. Neither had any experience with the use of any kind of response prompting procedures. Prerequisite skills needed to take part in the study were being able to pay attention to an activity for at least 10 minutes, and being able to follow directions given by his teacher or any other trainer.

Faruk could perform basic self help skills such as toileting, dressing, undressing, and eating independently. He could manage many fine motor skills (holding a pencil, drawing a picture, coloring a given shape, etc.) when told to do so and gross motor skills (walking up and down the stairs, kicking a ball, etc.) independently. He could use two or more word sentences. Faruk could also initiate and complete a short conversation.

Orhan could perform basic self-help skills such as eating, toileting and dressing with physical plus verbal prompts or only with verbal prompts. He could also perform many fine motor skills (buttoning his shirt, coloring a given shape, etc.) again with physical plus verbal prompts or only with verbal prompts and gross motor skills (catching a ball with both hands, running after a friend, etc.) with verbal

prompts. Orhan could use one word sentences in order to tell his requests or express himself.

Trainer and Observer

The trainer (first author) was a research assistant in a doctoral program in special education and also a special education class teacher in the university unit for children with developmental disabilities. She had four-years experience teaching children with autism. She conducted all sessions of the study. Reliability data were collected by a research assistant who was also a doctoral student in special education and a special education class teacher in the same unit. She had also experience in using response prompting procedures during instruction in her class with her students with developmental disabilities. The observer was provided information about collecting reliability data before she started.

Settings

All sessions, except generalization sessions, were conducted in a classroom of the university unit. The classroom was the participants' everyday classroom. Sessions were conducted during the lunch and play break. The unit was a segregated unit for children with developmental disabilities. The classroom had four square student desks, six small student chairs, four cupboards, one teacher table, one teacher chair, a chalkboard, an observation mirror and a video camera for recording the study. The student and the trainer sat at one of the student desks facing each other and all sessions were conducted in a 1:1 environment. The generalization sessions across materials were also generalization sessions across settings. They were conducted in the TV-Video Room of the unit where the TV, video, and many other educational materials were kept. Generalization sessions across trainers were conducted in the intervention setting with the second author.

Materials

There were eight portrait pictures of relatives for each participant. Pictures were on 10 cm x 15 cm cards. For generalization sessions across both materials and settings, the relatives' full body and portrait appearances were shown on TV to participants. For generalization, families were given a handycam camera before the study was started and asked to record all the full body appearances and portrait appearances of relatives identified by the researchers. For generalization sessions across trainers, pictures with the full body appearance of the relatives were on 10 cm x 15 cm cards. All cards were laminated. Also, chips as reinforcement and a stopwatch were used to record the duration of each session.

Procedure

Experimental procedure consisted of full probe, daily probe, training, maintenance, and generalization sessions. All sessions were conducted between 10 a.m. and 11 a.m. every school day in a 1:1 environment. Two daily probe sessions and two training sessions were conducted every day.

Full probe sessions. Full probe sessions were conducted before, and after the criterion was met by the participants during the training sessions. The first full probe session was conducted in order to collect baseline performance data. During all the full probe sessions, sixteen trials were conducted in order to check every target stimuli (i.e., relative pictures). In each teaching set, four relative pictures took place and during each full probe session, each target stimuli was asked three times randomly and 24 trials were conducted. After a participant met criterion with the first teaching set, the second full probe session was conducted. Similarly, after meeting criterion with the second set, last full probe session was conducted.

Full probe sessions were conducted as follows: (a) the materials (e.g., relative pictures, chips, stopwatch) were placed on the material desk, (b) relative pictures of the first teaching set were placed on the desk, (c) an attentional cue (i.e., "Orhan, are you ready?") was provided, a task direction (i.e., "Orhan who is this?") was given, (d) a 4 s waiting interval was provided for the participant's response, (e) correct responses were rewarded verbally, and (f) incorrect responses were ignored by the trainer. The next trial was conducted 4 s after the previous one.

Correct responses of participants were rewarded verbally during full probe sessions (i.e. well done, good boy, etc.). Correct response was naming the relative shown on the picture. Incorrect responses were ignored by the trainer. Incorrect responses were naming the relative shown on the picture wrongly, naming the relative after the 4 s interval, or not naming at all in 4 s. For attending the full probe conditions and their cooperation, participants were provided tangible reinforcers.

Daily probe sessions. During simultaneous prompting a controlling prompt is always provided in the trials. Therefore participants do not have the chance to respond independently to the task directions. For this reason daily probe sessions were conducted before each training session in order to give the chance to respond independently. Only before the first training session was there no daily probe sessions. During every daily probe sessions each target stimuli was asked three times and 12 trails were conducted.

The only difference between full probe sessions and daily probe sessions was that, during full probe sessions all target behaviors were asked of the participant, during daily probe sessions, only the target behavior being taught was asked. Task directions were presented to the participant in an unpredictable order.

As during full probe sessions, correct responses resulted with verbal praise and all incorrect responses were ignored. Also, attention and cooperation of participants were reinforced at the end of each daily probe session.

Simultaneous prompting training sessions. Simultaneous prompting procedure was conducted for teaching relative names. During training sessions, task direction and controlling prompt were delivered at the same time (0 s). In the training sessions each relative name was asked three times. As there were four relative names in each teaching set, 12 trials were conducted during each training session. The response and inter-trial intervals were 4 s during training sessions. Model prompting plus verbal prompting were used as controlling prompts. Training sessions were continued until three consecutive 90-100% correct responses were given during daily probe sessions.

Training sessions were conducted as follows: (a) materials (four relative portrait pictures) were prepared on the desk, (b) introduction of the study took place (i.e., "Now we are going to learn how to name our relatives. First I am going to say and you will listen. Then I am going to ask you and you will tell the name of the relative on the picture."), (c) an attentional cue was given to the participant (i.e., "Orhan, are you ready to work?"), (d) task direction was given (i.e., "Orhan, tell me, who is this?"), (e) controlling prompt and modeling were delivered (i.e., "Orhan, look this is aunt. Now you tell me, who is this?"), (f) participant's response interval of 4 s, (g) correct responses (i.e., "This is aunt.") resulted with verbal praise (i.e., "Good boy", "Bravo", etc.), (h) incorrect responses (were defined the same as in probe sessions) or no responses within 4 s resulted with error correction (i.e., "This is aunt."), (i) correct responses after error correction were also praised verbally by the trainer, (j) 4 s waiting interval was conducted before the next trial. At the end of each training session, attention and cooperation of the participants were reinforced with tangible reinforcers.

Maintenance and generalization probe sessions. Maintenance sessions were conducted one, two and four weeks after the last full probe session was conducted with each participant. Maintenance sessions were conducted the same as full probe sessions. During maintenance sessions participants' attention and cooperation behaviors were reinforced with tangible reinforcers.

Generalization data were collected across materials, settings, and across trainers. All generalization data were examined via a pre- and post-test design. Pre-test session was conducted after the first full probe session and post-test session was conducted after each participant met the criterion. In each generalization session, 24 trials took place, there were eight target behaviors in each teaching set and each set was asked three times. During generalization sessions, correct responses of participants were praised verbally continuously and incorrect responses were ignored by the trainer. As in the other probe sessions, participants were reinforced with tangible reinforcers for their attention and cooperation.

Experimental Design

A multiple probe design across behaviors was used to examine the effectiveness of simultaneous prompting in teaching relative names to children with autism and was replicated across two participants. The dependent measure was the percent of correctly naming relatives. The independent variable was simultaneous prompting procedure. The independent variable was introduced to teach naming the relatives shown by the trainer (Target behaviors identified for each participant are presented in Table 1).

Reliability

Reliability data were collected in at least 20% of all the sessions. Two kinds of reliability data were collected: dependent and independent variable reliabilities. Both reliability data were collected from the same sessions selected randomly. Dependent variable (inter observer) reliability data were calculated by number of agreements divided by number of agreements plus disagreements multiplied by 100 (Tawney & Gast, 1984; Tekin & Kircaali-Iftar, 2001). Independent variable (procedural) reliability was calculated by dividing number of teacher behaviors observed by number of teacher behaviors planned multiplied by 100 (Billingsley, White, & Munson, 1980; Tekin & Kircaali-Iftar). Teacher behaviors observed were as follows: (1) controlling materials, (2) securing attention, (3) delivering task direction, (4) delivering controlling prompt (for only training sessions), (5) waiting for the 4 s response interval, (6) giving appropriate responses for the participants' responses (error correction was conducted during training), and (7) waiting for intertrial interval.

TABLE 1

Target Behavior Sets

Students	*Sets*	*Naming Relatives*
Faruk	1	Babaanne (Father's mother)
		Anne (Mother)
		Anneanne (Mother's mother)
		Hala (Father's sister)
	2	Dayi (Mother's brother)
		Dede (Grandfather)
		Amca (Father's brother)
		Eniste (Aunt's husband)
Orhan	1	Anne (Mother)
		Baba (Father)
		Buyukbaba (Grandfather)
		Dayi (Mother's brother)
	2	Amca (Father's brother)
		Babaanne (Father's mother)
		Anneanne (Mother's mother)
		Dede (Grandfather)

Social Validity

Social validity data were collected via a questionnaire sent in closed envelopes to parents of participants. Five questions were asked in the questionnaires. These were: (a) Do you think that it is important to teach relative names to your child? (b) What are the important parts of the study we've conducted with your child? (c) Are there any parts of the study that you did not like? If yes, please indicate. (d) Are there any differences in your child after the study was completed? If yes, please indicate. and (e) Do you think there are any problems about the participation of your child in a study similar to the study we've conducted? Can you tell the reason for your answer in a few sentences? Parents wrote answers to the questions and sent the closed envelopes back to the researchers.

Results

Instructional Data

Results indicated both participants acquired the target behaviors. Simultaneous prompting was an effective way of teaching relative names to children with autism in this study. Figures 1 and 2 show percent of correct responses of participants during full probe, training, and maintenance sessions.

Table 2 presents each participant's number of training sessions, and trials, number of training errors, percent of training errors, length of training sessions, length of daily probe sessions, number of probe errors and percent of probe errors.

In Table 2, number of training sessions was 27, and number of trials was 324. It can be

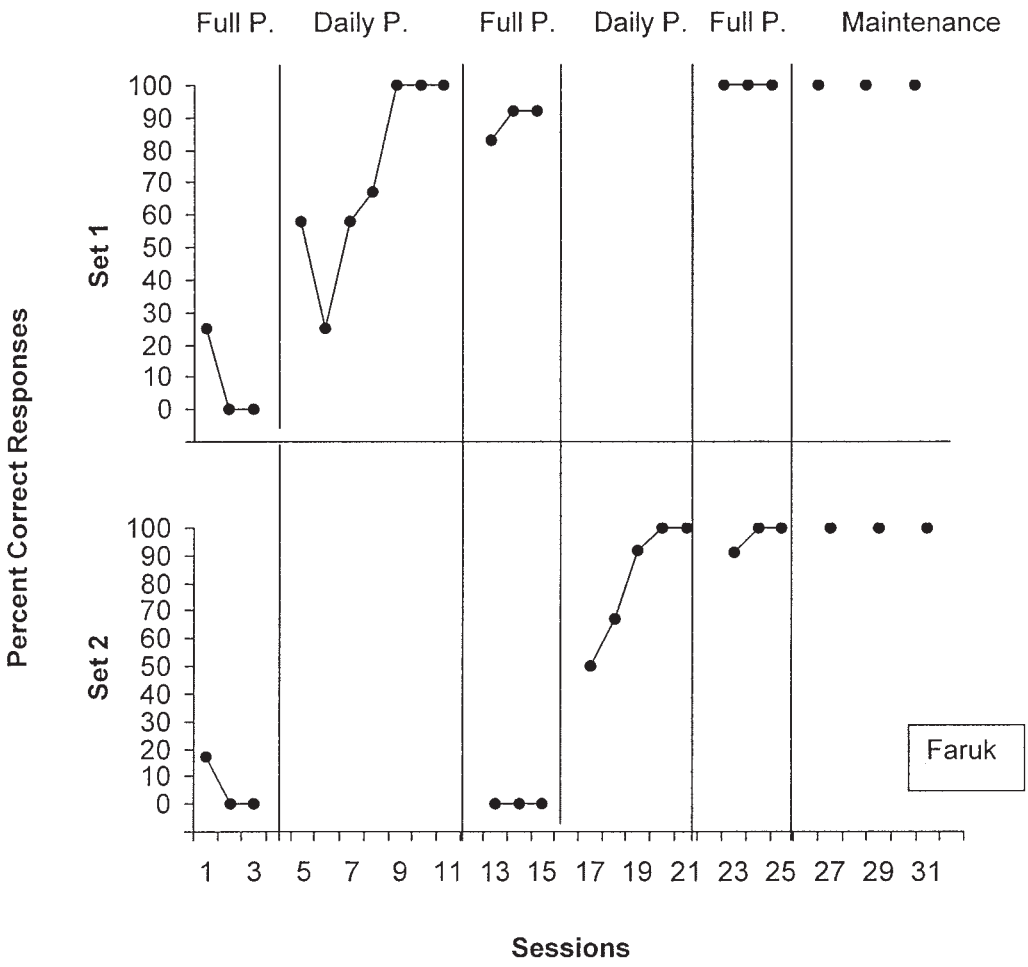

Figure 1. Percent correct responses for Faruk during full probe, daily probe, and maintenance probe sessions.

seen that length of all training sessions with two participants was 52 min and 58 s and length of daily probe sessions for two participants was 51 min and 9 s.

Maintenance and Generalization Data

Maintenance data show that both students maintained the skills taught during instruction one, two and four weeks after they met the criterion. Figures 1 and 2 show participants maintained the skills acquired during training sessions. Orhan maintained the skills taught 96.6% (range = 90-100) and Faruk maintained the skills taught 100% during the maintenance sessions.

Generalization data were collected across materials, across settings, and across trainers. For generalization across materials, relatives' full body appearances and portrait appearances were shown on the TV to the participants. During pre-test of the full body appearances on TV, Faruk responded with 8% accuracy (2 correct responses and 22 incorrect responses) and Orhan responded with 0% accuracy. During the post-test of the all body appearances on TV, Faruk responded with 75% accuracy (18 correct responses and 6 incorrect responses) and Orhan responded with 87.5% accuracy (21 correct responses and 3 incorrect responses).

For generalization across materials, but with

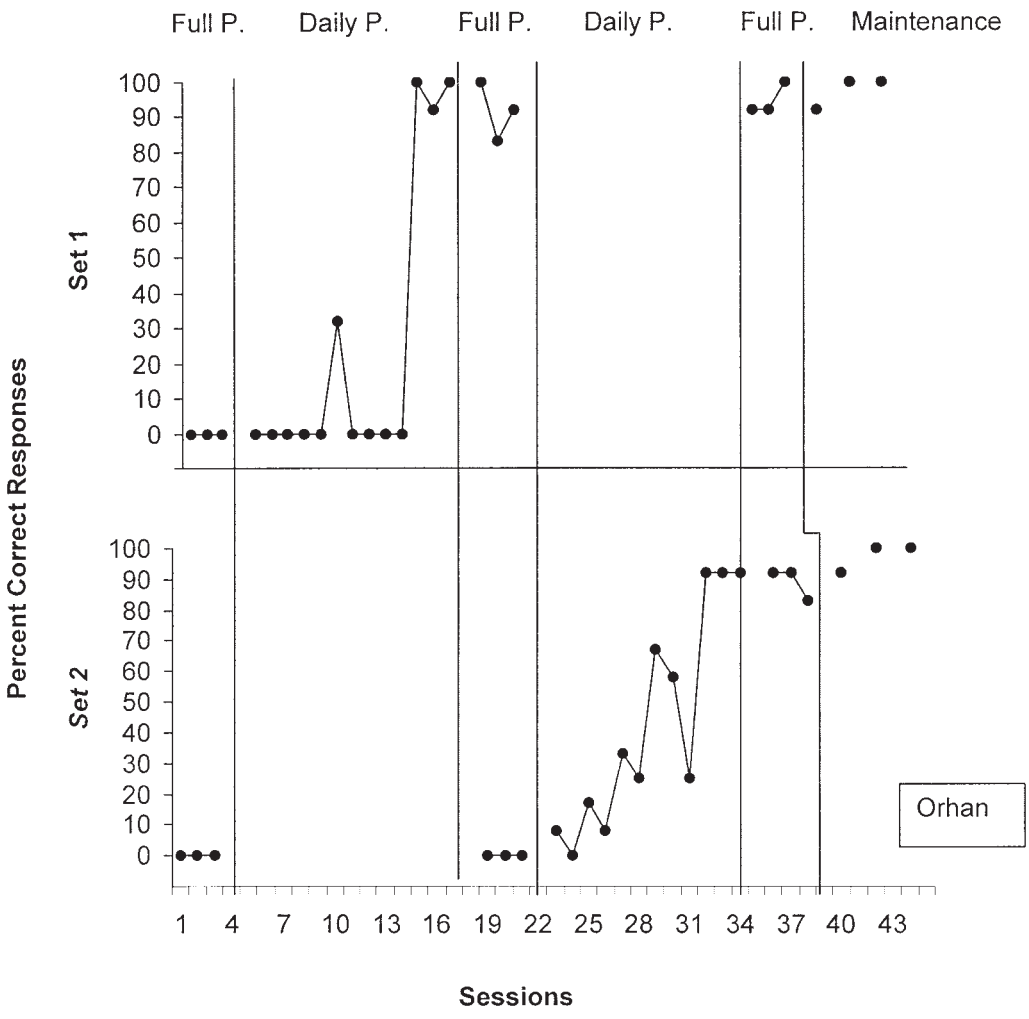

Figure 2. Percent correct responses for Orhan during full probe, daily probe, and maintenance probe sessions.

portrait appearances on the TV, the pre-test results were as follows: Faruk had 4% accuracy (1 correct response and 23 incorrect responses), and Orhan had 0% accuracy. During the post-test of the portrait appearances on TV, Faruk responded with 75% accuracy (18 correct responses and 6 incorrect responses), and Orhan responded with 83% accuracy (20 correct responses and 4 incorrect responses).

Generalization across settings sessions was conducted in the TV-Video Room of the unit. This room was in the same unit as the participants' class. In the room there were different materials being used by the classroom teachers during school days. Also there was a table and three chairs for watching the TV and video. The trainer conducted the generalization across settings sessions in a 1:1 environment. For generalization across settings, the relatives' full body appearances and portrait appearances were shown on the TV. Results of this part of generalization across settings were the same as the results of generalization across materials with full body appearances on TV.

For generalization across settings, but with portrait appearances on the TV, results of the application were the same as the generaliza-

TABLE 2

Evaluating Teaching Data and Teaching Sets

Students	*Set*	*No Training Sessions*	*No Training Trials*	*No Training Errors*	*% Training Errors*	*Training Time*	*Daily Probe Time*	*No. Probe Errors*	*% Probe Error*
Faruk	1	5	60	6	6,3	10 min 35 s.	10 min 3 s	21	25
	2	3	36	0	0	5 min 46 s	6 min 59 s	11	18.3
Total		8	96	6	3,1	16 min 21 s	17 min 2 s	32	21.6
Orhan	1	9	108	38	35	20 min 21 s	16 min 4 s	105	80
	2	10	120	7	0.5	16 min 16 s	17 min 45 s	83	57
Total		19	228	45	17.7	36 min 37 s	33 min 49 s	188	68.5
Grand Total		27	324	51	10.4	52 min 58 s	51 min 9 s	220	45

tion sessions across materials with portrait appearances on TV. Therefore detailed pre- and post-test results will not be presented again.

Generalization across trainers data were collected with the second author. During pre-test both participants responded with 0% accuracy, and during post-test, both participants responded with 100% accuracy.

Reliability Data

Dependent variable reliability (inter-observer reliability) data indicated 96.5% (range = 91.6 100%) agreement during full probe, training and maintenance and generalization sessions for Faruk. For Orhan there was an agreement of 98.2% (91.6-100%) during the full probe, training and maintenance and generalization sessions.

Results of independent variable reliability (procedural reliability) revealed that the trainer implemented the planned steps with 100% accuracy for both participants.

Social Validity Data

Both parents indicated that they were very happy because of their children being able to name their relatives at home. They also mentioned that the study was important because of their children's self confidence development. Also that their children started to use their relative names when they needed to communicate with them at home. Another point that parents mentioned was the development of using different words during communication after the study was completed. Besides all the points above, they also mentioned that they would give permission for their children's participation in similar studies.

Discussion

Results revealed that simultaneous prompting was an effective way of teaching relative names to children with autism. Also, results showed that participants maintained the skills taught one, two and four weeks after the training was completed, and generalized the skills across materials, settings, and trainers. These findings were consistent with some other studies examining the effectiveness of simultaneous prompting on discrete behaviors (Akmanoglu & Batu, 2004; Dogan & Tekin-Iftar, 2002; Gibson & Schuster, 1992; Griffen et al., 1998; MacFarland-Smith et al., 1993; Schuster, et al., 1992; Singleton et al., 1995).

During the study, the trainer provided reinforcement during daily and full probe sessions and also training sessions. This was implemented to show that the learning of participants was a result of the independent variable of the study.

Error correction was also provided to the participants during the study in the training sessions. Doing this, it was thought that the participants would learn the skills taught more quickly.

In this study, generalization data were collected across a number of variables. One of them was materials. Researchers used different kinds of photos (all body appearances and portrait appearances on TV) for the subjects to generalize the skills they acquired. Another

variable was across settings. During the generalization sessions participants were taken to the TV-Video Room of the unit. Also generalization across trainers data were collected during the study. Doing this, the researchers assumed that participants would generalize the skills they acquired more easily to their daily lives.

Teaching the relative names may have been easier because the trainer was the classroom teacher of the students. Therefore participants were very familiar with the trainer and were used to her instruction. Hence the researchers did not need a time for getting used to the participants during the study.

Some suggestions can be presented for the future studies. First, researchers can use simultaneous prompting for teaching both discrete and chained skills to individuals with autism and other developmental disabilities. Second, comparison of different errorless teaching techniques can be conducted for teaching various skills to individuals with or without disabilities.

References

Akmanoglu, N., & Batu, S. (2004). Teaching pointing to numerals to individuals with autism. *Education and Training in Developmental Disabilities, 39,* 326–336.

Billingsley, F., White, O. R., & Munson, R. (1980). Procedural reliability: A rationale and an example. *Behavioral Assessment, 2,* 229–241.

Coyne, P., Nyberg, C., & Vandenburg, M. L. (1999). *Developing leisure time skills for persons with autism.* Arlington: Future Horizons, Inc.

Dogan, O. S., & Tekin-Iftar, E. (2002). The effects of simultaneous prompting on teaching receptively identifying occupations from picture cards. *Research in Developmental Disabilities, 23,* 237–252.

Fetko, K. S., Schuster, J. W., Harley, D. A., & Collins, B. C. (1999). Using simultaneous prompting to teach a chained vocational task to young adults with severe intellectual disabilities. *Education and Training in Mental Retardation and Developmental Disabilities, 34,* 318–329.

Frost, L., & Bondy, A. (2002). *The picture exchange communication system training manual (2nd ed.).* Newark, DE: Pyramid Educational Products, Inc.

Gibson, A. N., & Schuster, J. W. (1992). The use of simultaneous prompting for teaching expressive word recognition to preschool children. *Topics in Early Childhood Special Education, 12,* 247–267.

Griffen, A. K., Schuster, J. W., & Morse, T. E. (1998). The acquisition of instructive feedback: A comparison versus intermittent presentation schedules. *Education and Training in Mental Retardation and Developmental Disabilities, 33,* 42–61.

Hodgdon, L. A. (1999). *Solving behavior problems in autism.* Michigan: Quirk Roberts Publishing.

Johnson, P., Schuster, J. W., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6,* 437–458.

Kircaali-Iftar, G. (2003). *Otistik ozellikler gosteren cocuklara iletisim becerilerinin kazandirilmasi (Teaching communication skills to children with autistic disorders).* Istanbul, Turkey: YA-PA Yayinlari.

Koegel, R. L., & Koegel, L. K. (1999). *Teaching children with autism.* Baltimore: Brookes.

MacFarland-Smith, J., Schuster, J. W., & Stevens, K. B. (1993). Using simultaneous prompting to teach expressive object identification to preschoolers with developmental delays. *Journal of Early Intervention, 17,* 50–60.

Palmer, T., Collins, B. C., & Schuster, J. W. (1999). The use of a simultaneous prompting procedure to teach receptive manual sign identification to adults with disabilities. *Journal of Developmental and Physical Disabilities, 11,* 179–191.

Parrott, K. A., Schuster, J. W., Collins, B. C., & Gassaway, L. J. (2000). Simultaneous prompting and instructive feedback when teaching chained tasks. *Journal of Behavioral Education, 10,* 3–19.

Riesen, T., McDonnell, J., Johnson, J. W., Polychronis, S., & Jameson, M. (2003). A comparison of constant time delay and simultaneous prompting within embedded instruction in general education classes with students with moderate to severe disabilities. *Journal of Behavioral Education, 12,* 241–259.

Schuster, J. W., & Griffen, A. K. (1993). Teaching a chained task with a simultaneous prompting procedure. *Journal of Behavioral Education, 3,* 299–315.

Schuster, J. W., Griffen, A. K., & Wollery, M. (1992). Comparison of simultaneous prompting and constant time delay procedures in teaching sight words to elementary students with moderate mental retardation. *Journal of Behavioral Education, 2,* 305–325.

Sewell, T. J., Collins, B. C., Hemmeter, M. L., & Schuster, J. W. (1998). Using simultaneous prompting within an activity based format to teach dressing skills to preschoolers with developmental delays. *Journal of Early Intervention, 21,* 132–145.

Singleton, D. K., Schuster, J. W., & Ault, M. J. (1995). Simultaneous prompting in a small group instructional arrangement. *Education and Training*

in Mental Retardation and Developmental Disabilities, 30, 218–230.

Singleton, D. K., Schuster, J. W., Morse, T. E., & Collins, B. C. (1999). A comparison of antecedent prompt and test and simultaneous prompting procedures in teaching grocery words to adolescents with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 34,* 182–199.

Smith, M. J. (2001). *Teaching play skills to children with autistic spectrum disorder.* New York: DRL Books, Inc.

Tawney, W. J., & Gast, L. D. (1984). *Single subject research in special education.* Columbus: Merrill Publishing.

Tekin-Iftar, E. (2003). Effectiveness of peer delivered simultaneous prompting on teaching community signs to students with developmental disabilities. *Education and Training in Developmental Disabilities, 38,* 77–94.

Tekin, E., & Kircaali-Iftar, G. (2001). *Ozel egitimde yanlissiz ogretim yontemleri (Errorless teaching procedures in special education).* Ankara, Turkey: Nobel Yayinevi.

Wolery, M., Ault, M. J., & Doyle, P. M. (1992). *Teaching students with moderate and severe handicaps: Use of response prompting strategies.* New York: Longman.

Wolery, M., Holcombe, A., Werts, M. G., & Cipolloni, R. M. (1993). Effects of simultaneous prompting and instructive feedback. *Early Education and Development, 4,* 20–31.

Received: 18 February 2004
Initial Acceptance: 7 April 2004
Final Acceptance: 30 September 2004

Education and Training in Developmental Disabilities, 2005, 40(4), 411–423

The Teacher Performance Rate and Accuracy Scale (TPRA): Training as Evaluation

Denise E. Ross
Teachers College, Columbia University

Jessica Singer-Dudek
St. John's University

R. Douglas Greer
Teachers College, Columbia University

Abstract: The purpose of this paper is to introduce the Teacher Performance Rate and Accuracy Scale (TPRA) which is a method of direct teacher observation used in the teacher evaluation and training component of the Comprehensive Application of Behavior Analysis to Schooling (CABAS®) model of schooling. The TPRA builds on the concept of academic engaged time (a measure frequently employed during ecobehavioral assessment) by counting the presence or absence of learn units (interlocking three-term contingencies for both students and teachers) during instruction. Implementation procedures for the TPRA, its application for identification and analysis of instructional problems, and its use for training and ongoing evaluation of teachers are presented and discussed.

Evaluating teachers' instructional effectiveness and providing feedback are components of teacher training that have been used to improve both teachers' performances and students' learning (Andrejko, 1998; Howard & McColskey, 2001; Munby, 1999; Rauch & Whittaker, 1999; Smith, Harris, & Sammons, 2001). Approaches to teacher evaluation have included reviewing professional development plans (Holland & Adams, 2002), examining teacher work samples (Denner, Salzman, & Bangert, 2001), conducting peer reviews (Kumrow & Dahlen, 2002), and evaluating professional portfolios (Moore & Bond, 2002). Research suggests that such approaches have been most effective when they occurred regularly, were part of proactive professional development programs, were based on multiple measures, and resulted in information to help improve instruction (Protheroe, 2002).

Such methods of teacher observation and feedback may often include indirect measures of instructional effectiveness such as parental questionnaires, students' evaluations, and teachers' self-reports (Bahr & Bahr, 1997; Hersen & Bellack, 2002). While these measures can provide feedback to teachers, research suggests that indirect measures of classroom instruction do not always reflect actual changes in instructional effectiveness or students' learning, but rather the reporter's verbal description of an intervention's effectiveness (Hawkins, 1991; Poling, Methot, & LeSage, 1995). For example, Miller and Kelley (1994) found that although homework completion rates increased, parents reported dissatisfaction with their children's rates of homework completion following an intervention to increase homework completion rates. Similar incongruent relationships between teachers' reports and students' performances have also been identified in both science and music education (Maranzano, 2000; Moore, 2003).

Some researchers have suggested that the primary measure of instructional effectiveness should be objective measures and not information obtained from subjective evaluations (Hawkins, 1991; Schwartz & Baer, 1991). Direct observation is one method of teacher evaluation and training that may more accu-

Correspondence concerning this article should be addressed to Denise E. Ross, Health and Behavior Studies, Teachers College, Columbia University, 525 West 120th Street, Box 223, New York, NY 10027.

rately and objectively show actual changes in both teachers' behaviors and students' learning than the previously discussed indirect measures (Hawkins; Schwartz & Baer). Research on classroom environments has shown that direct observational scales can measure both teachers' and students' behaviors which can, subsequently, be manipulated to increase students' learning (Greenwood, Carta, Kamps, Terry, & Delquadri, 1994). For example, research on classroom environmental variables has shown direct positive correlations between improved student learning and teachers' manipulation of students' rates of: 1) opportunities-to-respond (Greenwood, Hart, Walker, & Risley, 1994), 2) active student responding (Heward, 1994), 3) academic engaged time (Fisher et al., 1980) and 4) contingent reinforcement (Madsen, Becker, & Thomas, 1968).

A common method of measuring relationships between teachers' behaviors and students' learning is ecobehavioral assessment. This type of assessment uses momentary time sampling to compare students' behaviors in the context of teachers' behaviors and other classroom variables and to determine which variables promote high rates of academic engaged time (Greenwood & Delquadri, 2002; Logan & Malone, 1998). Greenwood, Carta, et al. (1994) described a computerized ecobehavioral software system—the Ecobehavioral Assessment Scale (EBASS)—designed to record the presence or absence of a students' behaviors (i.e., academic and inappropriate classroom behaviors), a teachers' behaviors (i.e., teacher's position in the classroom), and classroom variables (i.e., physical arrangement) in inclusive, specialized, and preschool settings (Greenwood & Delquadri). EBASS has been used to accurately record data on observable classroom variables such as effects of time of day on students' behavior (Muyskens & Ysseldyke, 1998), time spent on reading for students with learning disabilities and emotional/behavioral disorders (Vaughn, Levy, Coleman, & Bos, 2002), amount of academic engaged time for students with disabilities in inclusive settings (McDonnell, Thorson, McQuivey, & Kiefer-O'Donnell, 1997; Wallace, Anderson, & Bartholomay, 2002), and variables that promote engagement during academic responding (Greenwood, Horton, & Utley., 2002).

Ysseldyke and Christenson (as cited in Christenson & Anderson, 2002) designed The Instructional Environment Scale (TIES-II) which is another model of ecobehavioral assessment used to collect direct observational data on individual students' classroom behaviors across four categories: planning, management, delivery, and monitoring/evaluation (Spicuzza et al., 2001). Compiled information is reported in percentages of time and can be used to determine relationships between environmental factors and an individual student's learning and to promote factors that increase learning (Quinn & McDougal, 1998). TIES-II has been used to compare significant differences between variables in classrooms of typical elementary school children and children at-risk for severe emotional disabilities (Lago-Delello, 1998; Montague & Rinaldi, 2001) and to show the differential effects of a math curriculum on the pre and post academic achievement of students with varying ability levels (Spicuzza et al., 2001).

Research suggests that while ecobehavioral scales have been useful for identifying classroom variables that contribute to students' learning, their usefulness and efficiency for changing teachers' classroom instruction is limited for at least three reasons (Greenwood et al., 2002). First, while the primary purpose of such analyses has been to measure academic engaged time (Greenwood, Carta, et al., 1994; Logan & Keefe, 1997), ecobehavioral assessments have been largely restricted to descriptive-correlational analyses and have not resulted in changes to teachers' instructional behaviors that affect academic achievement (Greenwood et al., 2002; Juniper Gardens Children's Project, 2001). Second, codes used during ecobehavioral observations can be complex. Greenwood and Delquadri (2002) noted that, to date, EBASS has been limited to observations of single students because of the observations' complexity. Finally, implementing ecobehavioral assessments can become costly. Greenwood et al. (2002) suggested that use of computerized software for ecobehavioral assessment was expensive and required resources that were not routinely available to school personnel.

The purpose of the present paper is to introduce the Teacher Performance Rate and Accuracy Scale (TPRA), a relatively simple

and efficient method of direct teacher observation and evaluation that has been used to change teachers' behaviors and students' learning by employing a measure that builds on the concept of academic engaged time (Greer, 1994, 2002; Greer, Keohane, & Healy, 2002; Ingham & Greer, 1992). Specifically, the TPRA directly measures three-term contingencies presented by teachers in order to assess functional interrelationships between teachers' behaviors, students' responses, and instructional stimuli—interrelationships which research has shown to comprise academic engaged time (Heward, 1994). The TPRA is an integral part of the teacher evaluation and training component of the Comprehensive Application of Behavior Analysis to Schooling (CABAS®), which is a behavior analysis systems approach to education. The TPRA was designed to reflect and increase components of teaching shown to be effective in the literature such as higher rates of students' correct responses, higher rates of opportunities to respond, lower rates of students' incorrect responses, and lower rates of teachers' instructional errors (Greer, 2002).

Research Background

The TPRA builds on the concept of measuring academic engaged time by counting the presence or absence of *learn units* (Albers & Greer, 1991; Emurian, 2004; Emurian, Hu, Wang, & Durham, 2000; Greer, 1994; Greer, 2002; Greer, McCorkle, & Williams, 1989; Greer et al., 2002; Ingham & Greer, 1992; McDonough & Greer, 1999; Selinske, Greer, & Lodhi, 1991). Table 1 provides an example of a learn unit. Learn units consist of three-term contingencies for students and interlocking three-term contingencies for teachers. In other words, learn units measure occurrence of antecedents, behaviors, and consequences for both teachers and students during instruction. Greer (2002) termed the measure *learn unit* since both teachers and students "learn" from the interaction in a symbiotic manner. Using the learn unit as a measure of teaching allows for isolation of teaching and learning from moment-to-moment as students respond to teachers' instructions and teachers respond, in turn, based on behaviors of their students. Using the learn unit as a measure permits simultaneous assessment of both teachers and students since learn units are measures of teachers' behaviors and responses to learn units are measures of students' behaviors.

Findings from behavior analysis, the TPRA, and the learn unit comprise the research background for the TPRA. First, according to Greer (1994), at least four bodies of research from 1960-1990 in behavior analysis have contributed to the research background for the TPRA. These include research bodies showing relationships between: 1) rates of contingent teacher approvals and disapprovals and corresponding decreases and increases in undesirable student behaviors (Heward, 1994), 2) conditioning and generalized stimulus control which suggested that previously non-preferred academic tasks such as reading and writing could be reinforced to levels that would maintain behavior in non-instructional settings (Greer, 1994), 3) increased correct student responding and students who received high numbers of opportunities-to-respond (Greenwood, Hart, et al., 1994), and 4)

TABLE 1

Example of Learn Unit with Interlocking Three-term Contingencies for Teacher and Student

Three-term Contingencies for Teacher	*Instructional Components*	*Three-term Contingency for Student*
First teacher antecedent	*Student attends to teacher*	
First teacher behavior	*Teachers says, "Count to 10"*	Student antecedent
First teacher consequence/ second teacher antecedent	*Student correctly counts to 10*	Student behavior
Second teacher behavior	*Teachers says, "Good job!"*	Student consequence
Second teacher consequence	*Teacher records data and learn unit is completed*	

applied behavior analysis and academic behaviors such as those found in the research on Direct Instruction which showed that applied behavior analysis was useful for academic tasks as well as classroom management (Becker, 1992).

Second, research on the TPRA showed relationships between TPRA observations and students' learning. Three studies found correlations between high numbers of supervisor-conducted TPRA observations and increased teacher productivity, contingent consequences during instruction, and learning of students with severe mental retardation attending CABAS® schools (Greer, McCorkle, & Williams, 1989; Lamm & Greer, 1991; Selinske, Greer, & Lodhi, 1991). Additionally, Ingham and Greer (1992) found generalized and higher functional relationships between TPRA scores and correct student responses when compared to nonspecific feedback (i.e., "Nice lesson") for teachers of students with mental retardation. Albers and Greer (1991) showed similar results for students in remedial mathematics classrooms and also found that high rates of correct academic behaviors for teachers and students were not different for vocal versus written teacher instructional presentations.

Third, research on the learn unit resulted in response definitions for TPRA components. Research suggested that antecedent presentations should be unambiguous (Albers & Greer, 1991; Ingham & Greer, 1992) and that students must actively respond during response opportunities (Greenwood, Hart, et al., 1994; Heward, 1994). Additionally, student learning increased more with the consequence component than with the opportunity-to-respond alone (that is, the antecedent and intraresponse time) (Albers & Greer; Greenwood, Hart, et al.; Greer, 1996). Research also showed that postcedents emitted in the form of correction or reinforcement operations were necessary for student learning (Albers & Greer; Ingham & Greer) and that students should attend to written discriminative stimuli presented during correction procedures (Hogin, 1996). Research on the number and rate of learn unit presentations suggested that faster and higher rates of learn units resulted in increased correct responses by students (Ingham & Greer; Linhart-Kelly & Greer, 1997) as well as higher numbers of correct responses and objectives attained by students (Greer et al., 1989; Ingham & Greer; Selinske et al., 1991) and as much as four to seven times more correct responding than during baseline (Albers & Greer; Greer et al., 1989; Ingham & Greer). Finally, recent research suggests that learn units are useful for measuring and changing instruction in both college classrooms (Bahadourian, 2000) and computer-assisted instruction (Emurian, 2004). Subsequent research has shown that learning problems can be pinpointed and corrected based on the components of the learn unit (Greer, 2002; Keohane, 1997).

Implementing the TPRA

TPRA observations are generally conducted by trained supervisors or mentor teachers, and can be used with varying class sizes (i.e., whole groups or individual students) by an observer whose use of the TPRA has been calibrated to 90% accuracy across multiple observational sessions with a trainer. Figure 1 and the procedure outlined below illustrate steps to complete a TPRA observation form for a single student. Table 2 describes steps to complete the TPRA for groups of students and for complex academic behaviors.

First, the teacher and supervisor identify the following instructional components: 1) a target student, 2) a target instructional program, 3) operational definitions of target behaviors for the student and corresponding correct and incorrect responses, 4) schedule(s) of reinforcement for the instructional session, 5) antecedents and postcedents, 6) instructional conditions under which behavior should occur, and 7) necessary prerequisite skills that the student should have before instruction is implemented. The supervisor also observes availability of related instructional materials such as data collection forms, items serving as reinforcers, writing utensils, and curricular supplies (i.e., textbooks, flashcards). Finally, the supervisor and the teacher review the student's graphs to determine 1) trends in student's responses to the target program prior to the pending observation, 2) consistency of plotted data with accurate graphing protocol, and 3) appropriateness of the target program and expected level of student response.

Teacher Performance Rate and Accuracy Scale

Date: 1/14/03 School: Kennedy Middle School

Teacher: L. Wilson Observer: K. Smith

Student: T. Washington Program: Reading Letters A-C

	Teacher Antecedent	Student Behavior	Teacher Consequence
1.	(√)	—	C
2.	√	—	C
3.	√	+	R
4.	√	+	R
5.	√	+	(R)
6.	√	—	C
7.	√	+	R
8.	√	+	R
9.	√	+	R
10.	√	+	R
Correct/ Incorrect:	9/1	7/3	9/1

Teacher Number Per Minute Correct: 8 correct learn units = 2.25 learn units/minute

Teacher Number Per Minute Incorrect: 2 correct learn units = .56 learn units/minute

Student Number Per Minute Correct: 7 correct learn units = 1.97 learn units/minute

Student Number Per Minute Incorrect: 3 incorrect learn units = .85 learn units/minute

Converted Time: 3.55 min Actual Time: 3 min 33 sec

Figure 1. Example of a completed TPRA form.

Second, the supervisor and, when appropriate, the teacher (i.e., in a fluency program) start a timer (i.e., a stopwatch) or record the analog time in minutes and seconds. Timing during TPRA observations begins with the teacher's presentation of the first antecedent within a set of learn units which comprise an instructional session. The timer is stopped only if an interruption occurs (i.e., a problem with another student or school-related disruptions) in which case the timer begins again when instruction resumes. The timer is for-

TABLE 2

Procedures for TPRA Observations Conducted with Groups and Complex Academic Behaviors

1. Record learn units for classroom management separately by using an "A" for correct social approvals and a "D" for correct social disapprovals administered by the teacher.
2. Use written work that incorporates vocal responses such as worksheets or computerized programs.
3. Indicate the beginning and end of observational periods with vertical lines drawn on written work (i.e., worksheets).
4. Record vocal learn units (V) separately from written learn units (W).
5. Count written learn units when the student has reviewed corrections and completed changes. This may not occur until the next class period.
6. Elapsed time is calculated by adding the time required to complete problems on the day of the observation to the time involved to complete the corrections.
7. Homework time is included by adding the number of minutes in a day that responses are reviewed and corrected by students to the actual instructional time.

mally stopped when the teacher delivers the final postcedent. Elapsed time is used to calculate rate of instruction and TPRA scores.

During the instructional session, the teacher records students' correct and incorrect responses to instruction based on the operational definition described before the lesson. Correct student responses occur when the student emits a behavior consistent with the operational definition within a specified intraresponse time (i.e., 5 seconds). Incorrect student responses occur when the student emits a behavior inconsistent with the operational definition, does not emit the correct behavior within the specified intraresponse time, omits a response, or emits a correct response after emitting an incorrect response (i.e., self-correction). Incorrect responses and response omissions are recorded as minuses (−) and correct responses are recorded as pluses (+). Using event recording, teachers record responses immediately after each postcedent by using pencil and paper method (on a pre-existing data collection sheet), mechanical counters, or a computerized data collection tool. When responses to textual passages are the target behavior, the teacher records correctly and incorrectly read words by marking on an identical version of the passage that is photocopied or covered with a transparent overlay.

Using the TPRA form (see Figure 1), the supervisor records the accuracy of each component of the learn unit based on information collected with the teacher before the observational session. First, the supervisor measures the teacher's presentation of instructional antecedents as correct or incorrect. Correct antecedents occur if the teacher's vocal and/or nonvocal antecedent stimulus was unambiguous, consistent with the lesson plan or script, and, in the case of curricular materials, the target stimuli were flawless (i.e., targeted stimulus features were salient). Correct antecedent presentations are recorded under the "antecedent" column (A) with a checkmark. Incorrect antecedent presentations are recorded under the same column with a circled checkmark.

Second, students' responses are recorded in the same manner described above for the teacher. That is, the observer marks correct student responses with a plus (+) and incorrect student responses with a minus (−). These data are compared with the teacher's collection of student data upon completion of the observation for interobserver agreement purposes.

Third, the teacher's presentation of postcedents is recorded. These are measured as correct and incorrect based on their contingent relationship to the student's responses. That is, teachers should perform reinforcement operations contingent upon correct student responses and correction procedures contingent upon incorrect student responses. A correct reinforcement operation is defined as immediate presentation of a stimulus that functions as a reinforcer for the target student on the schedule of reinforcement specified for the instructional session. An appropriate correction procedure is defined by the scripted program, school, or classroom protocol. In most cases, correction procedures include the teacher's presentation of the lesson's antecedent stimulus with an accompanying prompt or model for the target behavior. Corrections are not reinforced and the stu-

dent is required to emit the corrected response. Table 3 identifies various teacher postcedents and the corresponding TPRA codes used to record them.

During or after formal observation, the supervisor provides feedback to the teacher in one or more of three forms. Supervisors may provide oral feedback by reviewing correct and incorrect learn units with the teacher based on the TPRA form. Written feedback is provided by recording comments about specific learn units on the TPRA form and giving the form to the teacher to review. Supervisors may also stop TPRA observations and correct teachers' behaviors immediately—in which case the TPRA observation begins again as instruction resumes. When this happens, the supervisor may change an instructional procedure, model a correct procedure, verbally prompt the teacher to perform an instructional operation, or explain a procedure. Feedback is not provided during a session in which an interruption would be disruptive for a student. Teachers respond to TPRA feedback by demonstrating that they can identify instructional problems and problem-solve using a decision tree (Greer, 2002; Keohane, 1997) by asking the supervisor questions based on written comments. When feedback is given within an instructional session, teachers correct problem areas during subsequent learn unit presentations or lessons.

TABLE 3

Codes for TPRA Observations

Code	*Response Definition*
√	Antecedent presented correctly
(√)	Antecedent presented incorrectly
+	Correct student response
−	Incorrect student response
R	Reinforcement operation presented correctly
(R)	Reinforcement operation omitted or presented incorrectly
C	Correction operation presented correctly
(C)	Correction operation omitted or presented incorrectly
D	Social disapproval presented correctly
(D)	Social disapproval omitted or presented incorrectly
A	Social approval presented correctly
(A)	Social approval omitted or presented incorrectly
W	Written antecedent presentation
V	Vocal antecedent presentation

Formulas

The first calculation is rate of correct and incorrect responses for the student. First, all responses (including response omissions) are tallied and total number of incorrect responses is separated from total number of correct responses. Next, elapsed time is converted into minutes by dividing seconds by 60. One minute and 20 seconds (1:20), for example, would become 1.33 minutes. Then, number of correct and incorrect responses is each divided by the converted time and a rate of correct and incorrect responses is obtained.

The second calculation is rate of correct and incorrect responses for the teacher. Learn units are correct if both antecedents and postcedents were presented accurately during observation. Learn units are incorrect if either antecedent or postcedent were presented incorrectly during observation. Both correct and incorrect learn units are summed and divided separately by the converted time to obtain a rate of correct and incorrect responses for the teacher.

For example, during a lesson whose duration was 10 minutes, a teacher delivered 20 instructional antecedents accurately but ignored two correct responses which means that the teacher presented 18 correct learn units with 2 errors. Since the lesson's duration was 10 minutes, 18 would be divided by 10 for a TPRA score of 1.8 correct learn units presented per minute and .2 incorrect learn units presented per minute. That means that students were given opportunities to actively respond and to receive reinforcement and feedback for their responses approximately twice per minute.

Analysis of Data

Comparing TPRA scores is restricted to the same instructional programs by the same teacher with the same students because of

variations in students' learning histories, curricular objectives, and levels of teachers' expertise. For instance, some learn units contain multiple behaviors instead of a single student response; such is the case for programs that involve larger teacher antecedents, larger student responses, varying schedules of reinforcement, and different types of reinforcers. When this happens, regardless of number of student behaviors, it is the teacher's delivery of the consequence that defines the learn unit. A common example is when students learning to write complex essays have fewer learn units than students learning to write 25 spelling words because the placement of the teacher's consequence during instruction determines size of the learn unit (at the end of an essay versus at the end of each spelling word). Similarly, completing a page of mathematical problems would include multiple responses to a single antecedent. Likewise, citing a 10-digit telephone number may begin with a single response (i.e., the number 4) but, eventually, require 10 different numbers for a single correct response.

TPRA scores can be displayed graphically across time or summed and divided by the number of observations to obtain a mean accuracy TPRA score. Improved TPRA scores suggest the following: 1) shorter latent time periods between learn units to students which translates into greater amounts of instruction, 2) fluent teacher presentations, and 3) increased contingency-shaped behaviors instead of rule-governed behaviors (i.e., teachers who emit automatic behaviors instead of accessing procedures to instruct). For both teachers and students, accurate rates should increase, inaccurate rates should decrease, and changes in students' performances should be analogous to changes in teachers' performances.

TPRA scores can be used in school-wide summaries of data and for teacher-supervisor conferences conducted after observations. Specifically, when mean weekly scores for a single teacher are integrated into a school's TPRA data, both composite and individual TPRA data can be used to help identify learning objectives for students and teachers. Composite data help schools analyze a number of variables that CABAS® research has shown to be functionally related to accurate student and teacher performance including the number of supervisor observations completed (Greer et al., 1989; Ingham & Greer, 1992), supervisor expertise in solving instructional problems (Greer, 2002; Keohane, 1997), and setting learn unit targets for teacher performance (Albers & Greer, 1991). When teachers' incorrect performances are relatively low and stable, cumulative data reflecting the number of observations with and without errors are displayed.

Use for Instructional Decisions

When accurate and inaccurate data do not reflect divergent trends in performance (that is, ascending and descending trends, respectively) but teacher performance is errorless, other components of instruction are reviewed as possible sources for student learning problems by using an instructional analysis decision protocol and the learn unit context (i.e., motivational variables or learning history) (Greer, 2002; Keohane, 1997). Possible sources of the problem may be that the (a) student lacks the prerequisite skills to respond to the material being presented, (b) student lacks the topography or the response, (c) instructional materials are insufficient for acquiring stimulus control, or d) schedule of reinforcement is too thin. While ability to change some curricular problems does not lie in the TPRA observation itself (i.e., what to teach when a student lacks prerequisite skills), the observational procedure allows an observer or teacher to indicate that this is the area of the problem. Later, the information can be used in conjunction with a skilled teacher mentor to change a curricular problem. Changing a curricular problem is usually completed by using the decision protocol (Keohane, 1997) to analyze the learn unit content and to determine which of 200 research-based tactics (Greer) is likely to solve a particular instructional or learning problem. Table 4 lists common instructional errors, their associated TPRA components, and potential solutions. Case studies illustrating the application of TPRA data to instructional decisions are provided below.

Case Studies

Changing instructional errors. Teachers A and B were co-teachers in an inclusive first grade classroom where Janet, a 6-year old girl

TABLE 4

Common Instructional Errors Related to TPRA Components

TPRA Component	*Instructional Errors*	*Potential Solutions*
Antecedent	Vocal or written antecedent presented incorrectly or inconsistently	Review antecedents in program script or protocol
Antecedent	Target stimuli flawed or not salient	Make target stimulus dimensions salient
Antecedent	Opportunity-to-respond inconsistent or omitted	Use timer to prompt teacher to begin and end intraresponse time (i.e., five seconds)
Antecedent	Incorrect prompt presented or correct prompt presented inconsistently	Review prompt levels and fading procedures with teacher
Behavior	Student lacks prerequisite skills	Teach student prerequisite skills
Behavior	Inappropriate prompt level	Use different prompt level or error correction procedure
Behavior	Response topography is too complex	Task analyze target behavior and include prompts where needed
Behavior	Size of learn unit too large	Task analyze target behavior to change size of learn unit
Consequence	Student has no opportunity-to-respond consequence component	Prompt teacher to include opportunity-to-respond
Consequence	Reinforcers not potent or satiation	Vary reinforcers or perform establishing operation
Consequence	Reinforcers not delivered on appropriate schedule	Decrease or increase schedule of reinforcement
Consequence	Reinforcers not delivered immediately or contingently	Prompt teacher to deliver reinforcers before he/she records students' data
Number per minute	Correct responses per minute are too low	Increase teacher's automatic/contingency-shaped learn unit presentations

with autism, was learning to read sight words aloud. Teachers used the learn unit—that is, interlocking three-term contingencies for the teacher with the potential operant for the student—to teach Janet the target words. The teachers alternately presented each of four 20-learn unit instructional sessions. Janet only emitted a mean of 14.5 correct responses (range, 13 to 18) and did not obtain the instructional objective of 90% mastery. TPRA observations showed that Teacher A presented 6.51 correct learn units and 0 incorrect learn units per minute, but Teacher B presented 4.56 correct learn units and 1.95 incorrect learn units per minute. Teacher B's errors were in the omission of an opportunity-to-respond during correction operations. The supervisor modeled the appropriate correction operation, observed Teacher B again, and Janet achieved mastery criteria within the next two instructional sessions.

Increasing number of learn units emitted per minute. Teacher C was a reading teacher for 9th grade students with learning disabilities. She received TPRA observations on her implementation of a new behavioral reading program following training. During one 30-minute observation, she presented 64 complete learn units, but 104 additional antecedents to which students responded but were not reinforced or corrected. The supervisor showed Teacher C the TPRA scores and requested that she present a higher number of contingent correction and reinforcement operations during the next lesson. During a second 30-minute observation, Teacher C presented 102 complete learn units and only 52 antecedents without consequences.

Training student teachers. As part of the masters of arts program at Teachers College, groups of student teachers wrote instructional programs for children in CABAS® schools which were, subsequently, evaluated using the TPRA. One group wrote an instructional program to teach children with disabilities to emit conversations in three increasingly complex social situations. Children obtained mastery criteria in the simplest social situations but not in the most complex social situation. Since multiple TPRA data showed that student teachers correctly presented all antecedents and consequences in each social situation, the mentoring teacher could eliminate faulty antecedents and consequences as the source of learning difficulty and examine other areas of the learn unit as possible sources for the children's low number of correct responses (i.e., prerequisite skills or the program's generalization components).

Conclusion

The purpose of this paper was to describe the TPRA by presenting its research background, implementation procedures, formulas and data analysis, and instructional implications. The TPRA, an integral component of teacher training and evaluation in CABAS® schools, measures teacher-student interactions during instruction by assessing frequency of learn units. Learn units are defined as interlocking three-term contingencies (antecedents, behaviors, and consequences) for teachers and students. During timed instructional sessions conducted with various group sizes (i.e., ranging from single students to whole classes), trained observers use event recording to measure the accuracy of teachers' antecedent and postcedent presentations as well as the accuracy of students' responses. The accuracy of the learn unit's components and subsequent rate calculations of teachers' and students' correct and incorrect learn units per minute are used to remediate instructional errors associated with the learn unit—that is, the antecedents, behaviors, and consequences for both teachers and students.

As previously mentioned, observations from ecobehavioral assessment literature provide one basis for why the TPRA may be valuable for changing teacher behavior. Specifically, ecobehavioral assessment was intended to identify variables that promote academic engagement (e.g., instructional tasks or groups) (Greenwood et al., 2002). However, while information derived from ecobehavioral assessment has increased students' academic responding (Greenwood et al.), to date ecobehavioral assessments have mostly been descriptive-correlational and have not identified ways to promote engagement through altered classroom instruction (Greenwood et al.). What appears to be lacking in the research literature is a description of a simple measure that assesses teachers' instructional behaviors and that can be used to change them (Greenwood et al.). The TPRA is a relatively simple and efficient method of teacher observation that has been used to change teachers' instructional behaviors during training and evaluation.

When the TPRA was used weekly over one or more academic years as part of a teacher training and evaluation program, researchers found correlational and functional relationships between its use by a trained observer and teachers' and students' instructional responses, including increased numbers of: 1) instructional sessions taught, 2) learning objectives achieved, 3) correct student responses, 4) learn units per minute, 5) opportunities-to-respond, and 6) presentations of learn units during non-observational periods (Greer et al., 1989; Ingham & Greer, 1992; Lamm & Greer, 1991; Selinske et al., 1991). Based on these studies, a possible explanation for the procedure's effectiveness and why it may be a valuable tool for teacher evaluation and training is provided below.

The TPRA may be a useful tool for evaluating teaching because its primary measure, learn units, not only builds on opportunity-to-respond, but may also explain what is important about academic engaged time. Both academic engaged time and opportunity-to-respond have been shown to predict academic achievement – that is, higher amounts of time spent on tasks that promote academic success and higher numbers of opportunities-to-respond are correlated with student achievement and on-task behavior (Greenwood, Hart, et al., 1994). However, research suggests that learn units produce more student learning than opportunities-to-respond alone (Heward,

1994; Ingham & Greer, 1992) and result in relatively high amounts of feedback for teachers about student learning and in relatively high amounts of on-task behaviors for students (Heward). Academic engaged time, as a measure, uses time as a primary dimension, but does not account for the number of learn units that teachers present during instruction (Heward). If learn units are the point of contact between teachers and students (Heward) then the learn unit may be a basic measure of teaching (Greer, 1994, 2002). As such, the TPRA, which measures learn units within time, is a valuable tool for measuring and effecting behaviors of teachers.

There is still much to learn about measuring teacher-student interactions at the teacher level. Over the past 20 years, an estimated 300,000 TPRA observations have been completed across at least 20 schools involving more than 500 teachers. As part of teacher training and observation in eight CABAS® schools in the United States, Ireland, and England, more than 3.8 million data points are generated from TPRA observations and learn unit data derived from instructional sessions with teachers, parents, students, and supervisor. Based on our experiences with this procedure, a primary benefit of the TPRA is that this observational tool is a simple procedure that can provide in-class training for and evaluation of teachers without costly, out-of-class workshops or equipment (Ingham & Greer, 1992).

References

Albers, A. E., & Greer, R. D. (1991). Is the three-term contingency trial a predictor of effective instruction? *Journal of Behavioral Education, 1,* 337–354.

Andrejko, L. (1998). The case for the teacher portfolio. *Journal of Staff Development, Staff 19,* 45–48.

Bahadourian, A. (2000). *The effects of learn units on student performance in two college classrooms.* Unpublished doctoral dissertation. Teachers College, Columbia University.

Bahr, M. W., & Bahr, C. M. (1997). Educational assessment in the next millennium: Contributions of technology. *Preventing School Failure, 41,* 90–94.

Becker, W. (1992). Direct instruction: A twenty-year review. In R. West & L. Hammerlynck (Eds.), *Design for educational excellence: The legacy of B. F. Skinner* (pp. 71–112). Longmont, CO: Sopris West.

Christenson, S. L., & Anderson, A. R. (2002). Commentary: The centrality of the learningcontext for students' academic enabler skills. *The School Psychology Review, 31,* 378–393.

Denner, P. R., Salzman, S. A., & Bangert, A. W. (2001). Linking teacher assessment to student performance: A benchmarking, generalizability, and validity study of the use of teacher work samples. *Journal of Personnel Evaluation in Education, 15,* 287–307.

Emurian, H. H., Hu, X., Wang, J., & Durham, D. (2000). Learning JAVA: A programmed instruction approach using Applets. *Computers in Human Behavior, 16,* 395–422.

Emurian, H. H. (2004). A Programmed Instruction Tutoring System for Java": Consideration of Learning Performance and Software Self-Efficacy. *Computers in Human Behavior, 20,* 423–459.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 7–32). Washington, DC: U.S. Department of Education.

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

Greenwood, C. R., & Delquadri, J. (2002). Code for instructional structure and student academic response: CISSAR. In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioral assessment techniques* (pp. 120–122). New York: Percheron Press.

Greenwood, C., R., Hart, B., Walker, D. I., & Risley, T. (1994). The opportunity to respond and academic performance revisited: A behavioral theory of developmental retardation and its prevention. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshelman, et al. (Eds.), *Behavior analysis in education: Focus on measurably superior education* (pp. 213–224). Pacific Grove, CA: Brooks/Cole.

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *The School Psychology Review, 31,* 328–349.

Greer, R. D. (1994). The measure of a teacher. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshelman, et al. (Eds.), *Behavior analysis in education: Focus on measurably superior education* (pp. 161–171). Pacific Grove, CA: Brooks/Cole.

Greer, R. D. (1996). The education crisis. In M. A. Mattaini & B. A. Thyer (Eds.), *Finding solutions to*

social problems: Behavioral strategies for change (pp. 113–146). Washington, DC: American Psychological Association.

Greer, R. D. (2002). *Designing teaching strategies: An applied behavior analysis systems approach.* New York: Academic Press.

Greer, R. D., McCorkle, N. P., & Williams, G. (1989). A sustained analysis of the behaviors of schooling. *Behavioral Residential Treatment, 4,* 113–141.

Greer, R. D., Keohane, D., & Healy, O. (2002). Quality and comprehensive applications of behavior analysis to schooling. *The Behavior Analyst Today, 3, 120–132.* Retrieved December 22, 2003, from http://www.behavior-analyst-online.org/BAT/BAT-32.pdf

Hawkins, R. P. (1991). Is social validity what we are interested in? Argument for a functional approach. *Journal of Applied Behavior Analysis, 24,* 205–213.

Hersen, M., & Bellack, A. S. (2002). *Dictionary of behavioral assessment techniques.* New York: Percheron Press.

Heward, W. L. (1994). Three "low-tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshelman, et al. (Eds.), *Behavior analysis in education: Focus on measurably superior education* (pp. 283–320). Pacific Grove, CA: Brooks/Cole.

Hogin, S. E. (1996). *Essential contingencies in correction procedures for increased learning in the context of the learn unit.* Unpublished doctoral dissertation. Columbia University, New York.

Holland, P. E., & Adams, P. (2002). Through the horns of a dilemma between instructional supervision and the summative evaluation of teaching. *International Journal of Leadership in Education, 5,* 227–247.

Howard, B. B., & McColskey, W. H. (2001). Evaluating experienced teachers. *Educational Leadership, 58,* 48–51.

Ingham, P., & Greer, R. D. (1992). Changes in student and teacher response in observed and generalized settings as a function of supervisor observations. *Journal of Applied Behavior Analysis, 25,* 153–164.

Juniper Gardens Children's Project. (2001). *EBASS.* Retrieved December 22, 2003, from University of Kansas, Juniper Gardens Children's Project Web site: http://www.jgcp.ku.edu/EBASS/ebass_descrp.htm

Keohane, D. (1997). *A functional relationship between teachers' use of scientific rule-governed strategies and student learning.* Unpublished doctoral dissertation, Columbia University, New York.

Kumrow, D., & Dahlen, B. (2002). Is peer review an effective approach for evaluating teachers? *Clearing House, 75,* 238–241.

Lago-Delello, E. (1998). Classroom dynamics and the development of serious emotional disturbance. *Exceptional Children, 64,* 479–492.

Lamm, N., & Greer, R. D. (1991). A systematic replication of CABAS in Italy. *Journal of Behavioral Education, 1,* 427–444.

Linhart-Kelly, R., & Greer, R. D. (1997, May). Rate mastery and the long-term maintenance of sight words. Paper presented at the 23rd Annual International Conference of the Association for Behavior Analysis, Chicago, IL.

Logan, K. R., & Keefe, E. B. (1997). A comparison of instructional context, teacher behavior, and engaged behavior for students with severe disabilities in general education and self-contained elementary classrooms. *The Journal of the Association for Persons with Severe Handicaps, 22,* 16–27.

Logan, K. R., & Malone, D. M. (1998). Instructional contexts for students with moderate, severe, and profound intellectual disabilities in general education elementary classrooms. *Education and Training in Mental Retardation and Developmental Disabilities, 33,* 62–75.

Madsen, C. H., Jr., Becker, W. C., & Thomas, D. R. (1968). Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis, 1,* 139–150.

Maranzano, C. (2000). Music teacher performance evaluation: A call for more inclusive models. *Journal of Personnel Evaluation in Education, 14,* 267–274.

McDonnell, J., Thorson, N., McQuivey, C., & Kiefer-O'Donnell, R. (1997). Academic engaged time of students with low-incidence disabilities in general education classrooms. *Mental Retardation, 35,* 18–26.

McDonough, S. H., & Greer, R. D. (1999). Is the learn unit a fundamental measure of pedagogy? *The Behavior Analyst, 22,* 5–16.

Miller, D. L., & Kelley, M. L. (1994). The use of goal setting and contingency contracting for improving children's homework performance. *Journal of Applied Behavior Analysis, 27,* 73–84.

Montague, M., & Rinaldi, C. (2001). Classroom dynamics and children at-risk: A follow-up. *Learning Disability Quarterly, Learning Disability 24,* 73–84.

Moore, J. (2003). Evaluating teaching is important. *Journal of Chemical Education, 80,* 119. Retrieved October 23, 2003 from http://JchemEd.chem-.wisc.edu Moore, Z., & Bond, N. (2002). The use of portfolios for in-service teacher assessment: A case study of foreign language middle-school teachers in Texas. *Foreign Language Annals, 35,* 85–92.

Munby, H. (1999). Planning, implementing, and evaluating a field-based teacher education pro-

gram: An introduction. *Teacher Education Quarterly, 26,* 5–9.

Muyskens, P., & Ysseldyke, J. E. (1998). Student academic responding as a function of classroom ecology and time of day. *The Journal of Special Education, 31,* 411–424.

Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research.* New York: Plenum Press.

Protheroe, N. (2002). Improving instruction through teacher observation. *Principal, 82,* 48–51.

Quinn, K. P., & McDougal, J. L. (1998). A mile wide and a mile deep: Comprehensive interventions for children and youth with emotional and behavioral disorders and their families. *The School Psychology Review, 27,* 191–203.

Rauch, K., & Whittaker, C. R. (1999). Observation and feedback during student teaching: learning from peers. *Action in Teacher Education, 21,* 1999.

Schwartz, I. S., & Baer, D. M. (1991). Social validity assessments: Is current practice state of the art? *Journal of Applied Behavior Analysis, 24,* 189–204.

Selinske, J., Greer, R. D., & Lodhi, S. (1991). A functional analysis of the comprehensive application of behavior analysis to schooling. *Journal of Applied Behavior Analysis, 24,* 107–117.

Smith, P. L., Harris, C. M., & Sammons, L. (2001). Using multimedia portfolios to assess preservice teacher and P-12 student learning. *Action in Teacher Education, 22,* 28–39.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39,* 521–542.

Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education, 36,* 2–13.

Wallace, T., Anderson, A. R., & Bartholomay, T. (2002). An ecobehavioral examination of high school classrooms that include students with disabilities. *Exceptional Children, 68,* 345–359.

Received: 11 February 2004
Initial Acceptance: 4 April 2004
Final Acceptance: 10 October 2004

INDEX—VOLUME 40, 2005

AUTHORS

TITLES

Statement of Ownership, Management, and Circulation

(Act of August 12, 1970, Section 3685 Title 39, United States Code)

1. Title of publication: EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES. **2. Publication number:** 1547-0350. **3. Date of Filing:** September 29, 2005. **4. Frequency of issue:** Quarterly in March, June, September, and December. **5. No. of issues published annually:** 4. **6. Annual subscription price:** $30 domestic; $75 institution; $79.50 foreign. **7. Location of known office of publication:** The Council for Exceptional Children, 1110 North Glebe Road, Suite 300, Arlington, Virginia 22101-5704. **8. Location of headquarters of general business offices of the publishers:** The Council for Exceptional Children, 1110 North Glebe Road, Suite 300, Arlington, Virginia 22201-5704. **9. Names and addresses of publisher and editor:** Publisher—The Council for Exceptional Children, 1110 North Glebe Road, Suite 300, Arlington, Virginia 22201; Editor—Dr. Stanley H. Zucker, Special Education Program, Farmer 316, Arizona State University, Tempe, AZ 85287-2011. **10. Owner:** The Council for Exceptional Children, Division on Developmental Disabilities, 1110 North Glebe Road, Suite 300, Arlington, Virginia 22201-5704, no stockholders. **[11.** Not applicable.**]** **12. For completion by nonprofit organizations authorized to mail at special rates:** The purpose, function, and nonprofit status of this organization and the exempt status for Federal Income Tax purposes have not changed during the preceding 12 months. **13. Publication name:** EDUCATION AND TRAINING IN DEVELOPMENTAL DISABILITIES. **14. Issue date for circulation data below:** September 29, 2005. **15. Extent and nature of circulation:**

	Average no. copies ea. issue for past 12 mos.	*No. of copies of issue nearest filing date*
A. Total no. copies printed	5,485	5,373
B. Paid circulation		
1. Sales through dealers and carriers, street vendors, and counter sales	0	0
2. Mail subscription	4,904	4,850
C. Total paid circulation	4,904	4,850
D. Free distribution by mail; samples, complimentary, and other free copies	150	125
E. Free distribution outside the mail; carriers or other means	0	0
F. Total free distribution	150	125
G. Total distribution	5,054	4,975
H. Copies not distributed		
1. Office use, left-over, unaccounted, spoiled after printing	421	398
2. Returns from news agents	0	0
I. Total	5,485	5,373
J. Percent paid and/or requested circulation	97.03%	97.48%

16. This statement of ownership will be printed in the Vol. 40 No. 4 DEC 05 issue of this publication.

17. Signature and title:

Drew W. Allbritten Executive Director September 29, 2005