Research and Practice in the Schools

The Official Journal of the Texas Association of School Psychologists

Volume 5, Issue 1 ◆ May 2018

Texas Association of School Psychologists
A partner on the path to student success

Research and Practice in the Schools:
The Official Journal of the Texas Association of School Psychologists

Volume 5, Issue 1 May 2018

Publication Information:
Research and Practice in the Schools is a peer-reviewed, online journal published by the Texas Association of School Psychologists. ISSN: 2329-5783

Research and Practice in the Schools:
The Official Journal of the Texas Association of School Psychologists

Research and Practice in the Schools is a publication of the Texas Association of School Psychologists (TASP). It is an online, peer-reviewed journal that provides TASP members with access to current research that impacts the practice of school psychology. The primary purpose of *Research and Practice in the Schools* is to meet the needs of TASP members for information on research-based practices in the field of school psychology. To meet this need, the journal welcomes timely and original empirical research, theoretical or conceptual articles, test reviews, book reviews, and software reviews. Qualitative and case-study research designs will be considered as appropriate, in addition to more traditional quantitative designs. All submissions should clearly articulate implications for the practice of psychology in the schools.

Instructions for Authors

General Submission Guidelines

All manuscripts should be submitted in electronic form to either of the co-editors (jeremy.sullivan@utsa.edu or aeherna8@uiwtx.edu) as an email attachment. Manuscripts should be submitted in MS Word format and labeled with the manuscript's title.

It is assumed that any manuscript submitted for review is not being considered concurrently by another journal. Each submission must be accompanied by a statement that it has not been simultaneously submitted for publication elsewhere, and has not been previously published.

Authors are responsible for obtaining permission to reproduce copyrighted material from other sources. IRB approval should have been obtained and should be noted in all studies involving human subjects. Manuscripts and accompanying materials become the property of the publisher. Upon acceptance for publication, authors will be asked to sign a publication agreement granting TASP permission to publish the manuscript. The editors reserve the right to edit the manuscript as necessary for publication if accepted.

Submissions should be typed, double-spaced with margins of one inch. All articles should meet the requirements of the *APA Publication Manual, 6th ed.*, in terms of style, references, and citations. Pages should be numbered consecutively throughout the document. Illustrations should be provided as clean digital files in .pdf format with a resolution of 300 dpi or higher. Tables and figures may be embedded in the text. A short descriptive title should appear above each table with a clear legend and any footnotes below.

The Review Process

After receiving the original manuscript, it will be reviewed by the Editors and anonymously by two or more reviewers from the Editorial Board or individuals appointed on an *ad hoc* basis. Reviewers will judge manuscripts according to a specified set of criteria, based on the type of submission. Upon completion of the initial review process, feedback will be offered to the original (primary) author with either (a) a preliminary target date for publication; (b) a request for minor editing or changes and resubmission; (c) significant changes with an invitation for resubmission once these changes are made; or, (d) a decision that the submission does not meet the requirements of *Research and Practice in the Schools*.

Call for Special Issue Proposals

We invite proposals for special issues of the journal, with the goal of publishing one special issue each year in addition to the general issue. Special issues will include collections of papers related to some cohesive theme in the field of School Psychology, and will be edited by Guest Editors who will take the lead in soliciting contributions and coordinating the peer review process. In addition to special issues that focus on research and scholarship in School Psychology, we welcome special issues that cover important practical and applied issues in the field.

Special issue proposals should include a brief description of the theme to be covered by the issue, approximate number of articles to be included, qualifications and expertise of those who will serve as Guest Editors of the issue, and a plan for soliciting manuscripts and conducting the reviews. Proposals for special issues, and questions about the process, should be sent to jeremy.sullivan@utsa.edu.

Graduate Student Section

Ashley Doss, doctoral student at Stephen F. Austin State University, serves as Graduate Student Section Editor for the journal. The Graduate Student Section is devoted to publishing the work of graduate students, including research studies, comprehensive literature reviews on relevant topics, and reviews of books or psychological/educational tests published within the past two years. As with all submissions to the journal, graduate student manuscripts should highlight implications for practice in the schools. If you are a graduate student and you have questions about how you can best contribute to the journal (as an author, reviewer, or both), please email Ashley at: dossan2@jacks.sfasu.edu.

Please note: all manuscripts submitted to the Graduate Student Section must include either a faculty co-author or a faculty sponsor who provides the student with mentorship on the process of preparing and submitting their work for peer review. When submitting their manuscripts for review, student authors should include a cover letter verifying that their work has been vetted by a faculty co-author or sponsor.

Self-Archiving Policy

Authors retain the right to self-archive the final, accepted manuscript of their submission on their own websites or deposit this version of the manuscript in any repository, provided it is only made publically available one calendar year (12 months) after publication or later. The archived version should be the final typeset article as it appears in the online issue of the Journal and the archive should include the appropriate citation and link to the Journal issue in which it appeared.

Research and Practice in the Schools
2018, Vol. 5, No. 1, pp. 1-8

ISSN: 2329-5783

Article

Teaching Problem-Solving Skills to Individuals with Disabilities Through the Use of Matrix Training

Kellie McKee, Glen L. McCuller, and Ginger L. Kelso

Stephen F. Austin State University

The present study examines the efficiency of matrix training as a format to teach problem-solving skills to two children with intellectual disabilities. Also called recombinative generalization (Goldstein, 1983), matrix training provides a way to organize skills to train more efficiently. However, it has typically been used to promote language acquisition. A 4X4 matrix of materials related to preparing a meal and potential associated problems was designed. Participants were taught a subset of the skills using least-to-most prompting and errorless learning. A multiple baseline across participants design was used to assess the effects of training on generalized material and problem combinations. Following training on a subset of skills, both participants demonstrated untrained material/problem combinations. These findings support previous literature on the efficiency of matrix training as a method to promote generalization to untrained behaviors. In addition, these results add the functional skill of problem-solving to the matrix training literature.

Key words: Matrix training, intellectual disability, problem-solving, multiple baseline design

Introduction and Literature Review

The variety of skills required for independent living and employment for people with intellectual disabilities are almost infinite. Considering the variety of ways any one task can be adapted to various situations, methods to promote generalization are critical. For example, consider the task of making toast. Although a rather routine activity, much about this task may change from day to day: type of bread, the bread container, the toaster and its location, number of pieces to make, and topping are just a few examples. The ability to adapt performance based on the current context can be referred to as problem-solving. Hughes and Rusch (1989) defined problem-solving as generalizing performances across various situations, settings, and people that may be unfamiliar. However, the ability to engage in problem-solving is an area of deficit for individuals with intellectual and developmental disabilities.

Problem-Solving

Hughes and Rusch (1989) confirmed that individuals with intellectual disabilities can be taught to solve problems as well as generalize performance within a work setting. They taught problem-solving skills to individuals with severe intellectual disabilities through the use of self-instruction and training of multiple examples of problems that occur within the work setting. Participants were required to perform various tasks involved in boxing and packaging soap, selecting and wearing correct garments, and checking to see the tasks for which they would be responsible for each day. Participants were taught to verbally state the problem, the correct response, and a consequence such as "good job!" Results showed that the participants successfully solved problems that had been trained and also generalized to untrained problems.

Martella, Agran, and Marchand-Martella (1992) also investigated problem-solving skills within the work setting. This study targeted safety skills to prevent work related injuries. Problem-

solving was measured through interviews and staged generalization assessments. Participants were trained to solve problems through the use of cue cards. These cards prompted the individual to verbalize four aspects of problem-solving: 1) How did the problem happen? 2) When could the problem be prevented? 3) Who could you talk to about the problem? and 4) What could be done or said in this situation? Following mastery of training, participants were required to continue using these steps to solve problems with corrective feedback, but without the support of cue cards. Results showed that participants improved their problem-solving abilities and generalized to similar and dissimilar situations.

Park and Gaylord-Ross (1989) focused on problem-solving for social skills in a work environment. They also used a four step problem-solving approach. The four steps were to decode the context of the social situation, decide on the correct behavior, perform the skill, and evaluate their performance. The researchers identified three target behaviors: lowering the occurrences of mumbling, initiation of social interactions, and expanding conversational skills. The four problem-solving steps were taught through role-playing. The participants were also taught seven rules to follow when attempting to solve problems. Results showed an increase in initiated conversations, decrease in mumbling, and increase in expansion across all three participants. This study resulted in generalization of social interaction but only had a lasting effect for one participant after training was implemented.

The previous articles promote various ways to teach problem-solving to individuals with disabilities. One theme found in Martella et. al (1992) and Park and Gaylord-Ross (1989) was that training for problem-solving occurred in multiple phases, which highlights the issue of efficiency. Implementing training that yields successful results is important. However, an equally important part of training for problem-solving is the issue of efficient methods to promote generalization to untrained problems. Another problem highlighted in the Park and Gaylord-Ross (1989) study was lasting effects of the training implemented. Also, in Hughes and Rusch (1989), training consisted of primarily teaching participants to verbalize problem-solving tasks but did not have the participant physically problem-solve until training was withdrawn. Efficiency is an important concern in training problem-solving because the goal is not only to promote skills quickly but also to yield lasting results and generalized problem-solving.

Matrix Training

One approach to improve efficiency and promote generalization across skills is matrix training. Also called recombinative generalization (Goldstein, 1983), matrix training provides an organized approach to teach a subset of relations and promote generalization to other untrained relations. The relations are typically tabled, placed along x and y-axes, and cells are created to show which relations are trained and tested. For example, Figure 1 presents three nouns and three verbs, which combined can construct nine possible two-word sentences. By training the three cells on the diagonal labeled with a "T" the learner will have been exposed to all words. Subsequently, the untrained combinations may be demonstrated without additional training.

	Eat	Sleep	Jump
Dog	T		
Cat		T	
Rabbit			T

Figure 1. *A simplified example of a matrix used to train two-word sentences.*

Goldstein and Mousetis (1989) used matrix training to promote expressive language in children with intellectual disabilities using peer models. A 7 X 7 matrix was created for the three participants with words that described an object placed along the vertical axis. Location words were placed along a horizontal axis. Examples of two word combinations were "scale, desk" "comb, dresser" and "button, table". Participants would be asked to

place objects in the various locations. The participants or model peer were presented verbally with one object and location and were expected to explain to the trainer which location they should place the object. Next, the trainer would provide an instruction to the target participant in which the participant was expected to place the object in the proper location. The model peer only responded verbally and did not model the behavior of placing an object in a location. Targeted participants learned and generalized correct verbal response combinations and demonstrated them through the use of matrix training.

Axe and Sainato (2010) used matrix training to promote generalized literacy skills with preschoolers with autism. The four participants were also identified as having significant delays in instruction following, writing skills, picture identification, and labeling. A 6 X 6 matrix of action-object combinations was designed. Actions included responses such as underlining or stamping and objects were pictures such as a pepper or a deer. Two participants required only the minimum six relations on the diagonal for generalization to occur. The other two required training on additional cells to generalize to untrained combinations.

More recent research has replicated and extended previous findings on the usefulness of matrix training to teach action-object combinations to children with autism (Curiel, Sainato, & Goldstein, 2016). However, several other educational goals have been addressed through the use of matrix training in recent years. MacManus, MacDonald, and Ahearn (2015) taught children with autism to engage in pretend play skills using an intervention that combined video modeling and matrix training. This study resulted in increases in untrained play sequences being acted out for all three participants. However, the participants needed varying levels of exposure to videos to produce generalization across all of the dimensions of the play sequence. Tanji and Noro (2011) also used matrix training for children with autism, but they focused on training spelling in the Japanese language. Following matrix training, the children were then tested to see if they could correctly spell words that recombine syllables in new ways. Results show that one of the two children immediately began generalizing spelling to the novel words after training. The other child needed to be trained on additional words to produce generalized spelling.

While recent research addressed the use of matrix training in the population of children with autism, research also exists on the use of matrix training to teach skills to adults without disabilities. Mahon, Lyddy, and Barnes-Holmes (2010) taught adults to respond to sound-symbol combinations by choosing the correct sound when presented with a symbol or choosing the correct symbol when presented with a sound. Those individuals who had been directly taught a subset of sound-symbol combinations were able to recombine those sounds and symbols in new combinations. Perez and de Rose (2010) also used matrix training to teach an adult without disabilities to read music. The participant was trained to select the musical notation that corresponded with a two-note sequence presented as auditory sounds. After the participant could accurately respond to each two-note sequence, she was tested on whether she could choose the correct musical notation after hearing three and four-note sequences. The participant correctly responded to 11 out of 12 possible three and four-note sequences. This study shows that matrix training can be applied to the area of learning musical notation.

As discussed, matrix training is an efficient and successful method to teach a variety of skills including receptive and expressive language. However, no research is currently available examining the use of matrix training to teach problem-solving skills to individuals with developmental disabilities. Matrix training provides an organized way of training a specific set of relations and promoting generalization to untrained combinations. This method of training could expand on the existing literature to promote adaptive behavior by providing a more successful and efficient approach to train individuals with disabilities to problem-solve unfamiliar tasks. The purpose of this research was to determine if matrix training can be used to successfully and efficiently teach generalized problem-solving skills. The use of matrix training to teach problem-solving may provide educational professionals an efficient and effective method to increase functional skills for children with disabilities.

Method

Participants

Two high school students from a rural school district participated in this study. Special education teachers identified both students as lacking problem-solving skills when completing functional tasks such as cooking or laundry. When presented with problems that occur when completing functional tasks, the participants would often stop working until a teacher or aide came to help. Rob was an 18-year-old Caucasian male in a self-contained Life Skills classroom. Rob was diagnosed with Autism Spectrum Disorder and moderate Intellectual Disability. His Full Scale IQ score was 53, and the extent of his verbal abilities were a few words and phrases such as "yes", "more", "thank you". Blake was an 18-year-old African-American male enrolled in the same self-contained Life Skills class. Blake had a diagnosis of Autism Spectrum Disorder and Moderate Intellectual Disability and a Nonverbal IQ score of 47. Blake communicated through gestures, facial expressions, and some sign language. Data concerning the specific intelligence test administered for each participant is unavailable. This presents a limitation in participant description. Both Rob and Blake were residents at a supported-living residential center and shared a dormitory with each other. All procedures were approved by the Institutional Review Board, and the guardians of both participants returned signed informed consent forms.

Setting

The study took place in the Life Skills classroom in which the two participants attended each school day. The classroom consisted of an area designated for computers, washer and dryer for laundry, desks, and a kitchen. The kitchen area had a stove, drawers and cabinets, a sink, and a refrigerator. All testing and training took place in the kitchen area of the classroom.

Materials

Materials used for testing and training related to preparing to cook pancakes and included pancake mix, a spatula, a wooden mixing spoon, a large size silver bowl, and a one-cup size measuring cup. Participants were expected to gather all the items from various places in the kitchen to prepare for making pancakes. The pancake mix was already placed out prior to the participants gathering the other materials needed.

	Bowl	Spoon	Measuring Cup	Spatula
Dirty	T			
Broken		T		
Wrong Size			T	
Missing				T

Figure 2. *Matrix for Training Materials to Problems.*
Note: Cells labeled with "T" represent paired relations used for training sessions.

General Procedures

Participants were given the task of gathering materials needed to make pancakes. Participants were not required to cook the pancakes, only gather materials. While gathering the four items needed to prepare pancakes, four different problems could be confronted with each item. The 4 X 4 matrix presented in Figure 2 illustrates the 16 potential material-problem combinations. The four cells labeled with a "T" represent the combinations trained.

During each session, assessment of problem-solving performance occurred prior to any training procedures. Upon entering the kitchen area, participants were asked to gather the four materials needed to make pancakes, and the trainer then recorded how many problems the participant solved correctly. During training sessions, the trainer presented the participant with one problem at a time. The trainer then asked the participant to show them how to correct the problem, and if the participant did not correctly solve the problem, least-to-most prompting and errorless learning was implemented.

Dependent Variable and Recording Procedures

The dependent measure was the number of problems and material combinations that the participant solved correctly. For example, if a bowl was broken, correcting the problem would mean finding an unbroken bowl. Materials consisted of (a) a large silver bowl, (b) a wooden spoon, (c) a spatula, and (d) a one-cup size measuring cup. The four problems identified were (a) dirty, (b) broken, (c) wrong size, and (d) missing. The participants were only asked to gather the materials, but were not asked to make the pancakes. Each testing session contained eight trials, four problem and four non-problem trials. The four problem trials consisted of one problem and material combination that had previously been trained. The other three problem trials were intended to assess for generalized problem-solving to untrained combinations. Non-problem trials were included to assess if the participant could gather needed materials without the need to problem-solve. Also, they were included to ensure the participants did not always "look" for a problem. The order of problem and non-problem trials were randomly assigned for each session. Over the course of four sessions, all 16 possible material and problem combinations were sampled.

Interobserver Agreement

Interobserver agreement was calculated by totaling the number of agreements between two observers during baseline, testing, and training sessions. IOA was taken by having a second observer record for both training and testing sessions at the same time as the first observer. Interobserver agreement was assessed across 65% of all sessions. IOA for sessions was calculated to be at 100% agreement (N=27).

Experimental Design and Conditions

A multiple baseline across participants design (Barlow, Nock, & Hersen, 2009) was used to assess the effects of training on generalized material and problem combinations. The multiple baseline design was used in order to demonstrate a functional relation between training a subset of problem-solving combinations and generalization to untrained combinations. The second participant remained in baseline until the first participant's data achieved stability. In this study, stability was defined as three consecutive sessions with two or more correct responses.

Baseline. All sessions took place in the kitchen area of the Life Skills classroom and typically lasted approximately 20 minutes. During each baseline trial, participants were told, "Today, we are going to get everything we need to make pancakes. You will need a silver bowl, a wooden spoon, a one-cup size measuring cup, and a spatula. Go ahead." No other instructions were given. No feedback or reinforcement was used contingent on performance during testing. Time to play on an iPad was provided at the conclusion of a session contingent on participation. The session contained eight trials with four problems and four non-problems. Participants entered training only after data in baseline remained stable.

Intervention. During intervention, participants began each session with testing using the same procedures followed in baseline. Immediately following testing, training began. Trials chosen for training were: a dirty bowl, a broken spoon, a wrong size-measuring cup, and a missing spatula. These trials were chosen from the cells on the matrix diagonal so that participants would sample each of the four materials and problem combinations (see Figure 2). The trainer followed a scripted three-step instruction sheet during each training session. First, she said to the participant "Today, we are going to learn how to deal with a problem." Second, the trainer would show and tell the participant which objects they were using and the problem. For example, when training dirty bowl, the trainer would then hold up an object and say, "This is a bowl. This bowl is dirty." Finally, the

participant was instructed to problem-solve by saying, "Show me what to do with a dirty bowl." The trainer allowed the participant 10 seconds to fix the problem. If the participant failed or did not attempt to problem-solve on the first attempt the trainer proceeded with least-to-most prompting until the problem was solved correctly. The prompts levels used were verbal, gestures, modeling, and physical guidance. The three steps followed by the trainer were altered somewhat for other items whereby participants would be shown where to find missing or correct sized items, or to replace broken items. Verbal praise was provided when the training task was completed correctly, even if prompts were needed. The criterion for ending training was if the participant completed the training without any prompting for a minimum of two training sessions. The participants were trained on four problems one after the other during each training session. Only one training attempt was completed for each problem during each session. The four problems trained remained the same throughout training. Following intervention sessions participants were given time to engage with an iPad for their participation.

Maintenance. Generalization and testing probes continued following participants' successful completion of training. These were presented the same as baseline and intervention sessions. Since time constraints prevented ongoing or long-term assessment of maintenance, generalization and testing probes were continued for ten sessions for Rob and for four sessions after training was completed for Blake to assess for continuation of trained and generalized skills.

Results

Figure 3 shows results for Rob and Blake for testing and generalization probes. Rob completed training criterion of requiring no prompts in three training sessions. On test probes of combinations included in training, Rob was correct on all probes after the first day of training. During generalization probes, Rob was able to generalize problem-solving skills to untrained problems in all combinations except for the second day of intervention where he demonstrated two of the three. Rob maintained generalization to untrained problems for a minimum of two consecutive days.

Blake completed training criterion in twelve sessions to achieve criterion of no prompts on training combinations for a minimum of two days. During training sessions three and four, Blake was noncompliant so training was not completed during those two sessions. On test probes of trained combinations he immediately began demonstrating problem-solving skills, except for session 12. On generalization probes of untrained combinations, Blake correctly performed all possible combinations, but his performance demonstrated variability. He was able to generalize to all possible combinations of materials and problems while training continued. Similarly, after achieving the training criterion his performance remained varied during maintenance probes.

Discussion

Following training on a subset of material/problem combinations related to meal preparation, both participants demonstrated generalization to untrained problem combinations when completing the task of gathering materials to make pancakes. The material/problem combinations were organized into a matrix and participants were trained on a subset, which sampled all possible materials and problems. Generalization probes were then implemented to assess for generalization to untrained combinations. Rob reached generalization to untrained problems in fewer sessions than Blake. Rob also met training criterion earlier during training. Blake also achieved generalization of untrained problems, but he required more training and his performance remained variable.

This study adds to the literature of efficient methods to teach problem-solving skills to individuals with intellectual disabilities. Importantly, the current study supports the findings of Goldstein and Mousetis (1989) in using matrix training as a method of teaching effectively and efficiently. In addition, these findings add problem-solving to the matrix training literature. That is, in addition to the skills identified by Goldstein (1983) such as language acquisition, problem-solving appears to be an appropriate target skill for matrix training methodology.

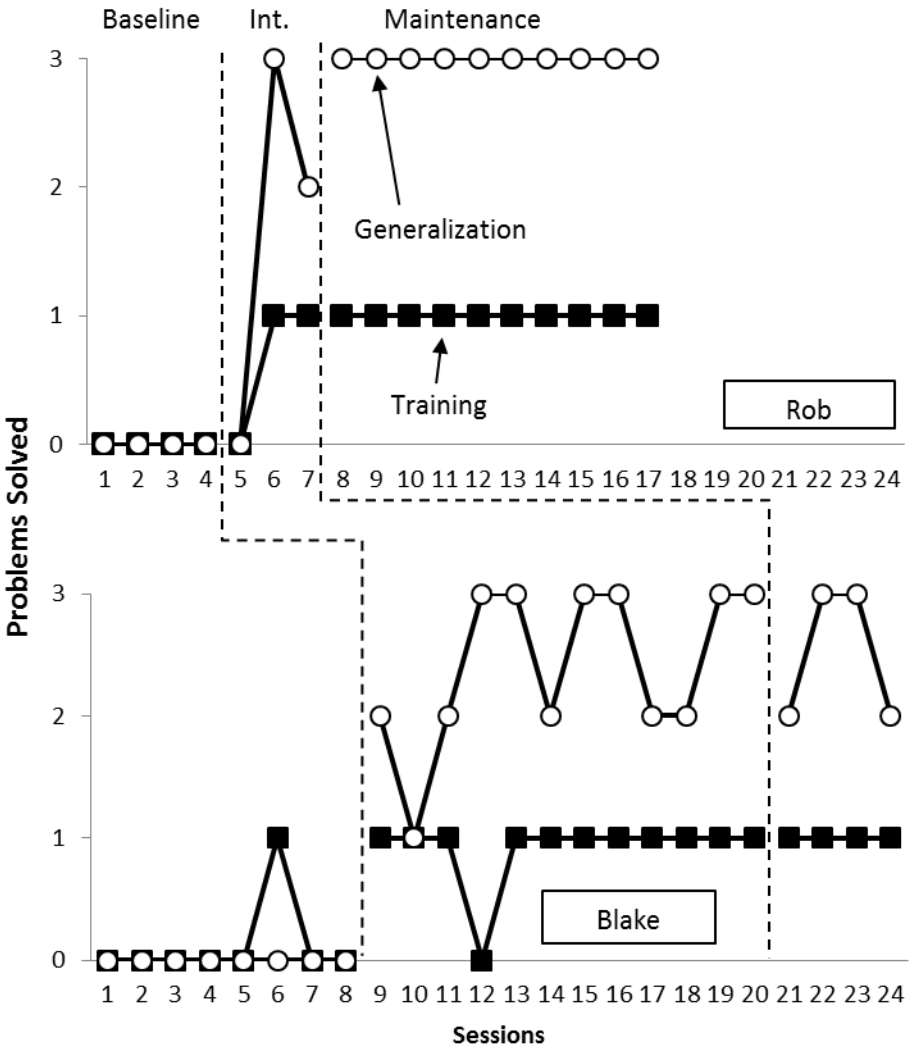

Figure 3. *Number of correctly solved problems on trained and generalized material and problem combinations.*

While the outcomes of this study support the use of matrix training for the instruction and generalization of problem-solving skills, there are limitations and recommendations to discuss. First, the study was limited to two participants. Due to the end of the school year and behavioral issues, the researcher was unable to gain more participants. Future studies may yield stronger results with more participants. Therefore, a functional relation could not be proven, but these results indicate a strong correlation between the independent and dependent variables. Single-subject research designs, such as the multiple-baseline across subjects design, are recommended to include three or more replications of effect to demonstrate experimental control (Horner et. al, 2005). Additional participants, including those with differing types and levels of disability, would enhance the external validity of using matrix training to address problem-solving. A second limitation is the possibility of a practice effect due to overexposure to the particular problems taught and assessed in this study. Future research should target skills with a broader range of materials and potential problems. By doing so, it may be possible to avoid overuse of examples across training and testing. This not only will address any potential practice effect, it would broaden the external validity of the results as well. A third limitation was that the researcher was not able to assess for long-term maintenance after

training had ceased due to the school year ending. It is important to note that after the study was completed, Rob's teacher informed the researcher of the fact that Rob was able and more willing to look for items in the kitchen if the class was performing the task of cooking. Given more time, future research could assess for maintenance of problem-solving by implementing maintenance probes after a longer period of time has passed. A final limitation to the present study is that it is unknown how well these skills would generalize across settings, materials, or other problem-solving situations. Further research could better assess generalization as it would be important to know if participants would demonstrate these learned skills in another kitchen, such as in a restaurant or home, using other examples of kitchen supplies, or with other tasks or situations in which problems may arise.

Matrix training promotes generalization through training enough examples of skills in order to promote generalization to other, nontrained skills (Stokes and Baer, 1977). This study suggests that matrix training has implications for organizing skills in order to efficiently teach problem-solving. For school psychologists, matrix training is a potentially useful tool for planning or teaching a variety of functional skills to individuals in need.

References

Axe, J. B., & Sainato, D. M., (2010) Matrix training of preliteracy skills with preschoolers with autism. *Journal of Applied Behavior Analysis, 43*, 635-652. doi: 10.1901/jaba.2010.43-635

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change*. Boston, MA: Pearson.

Curiel, E. S., Sainato, D. M., & Goldstein, H. (2016). Matrix training of receptive language skills with a toddler with autism spectrum disorder: A case study. *Education and Treatment of Children, 39*(1), 95-109.

Goldstein, H. (1983). Recombinative generalization: Relationships between environmental conditions and the linguistic repertoires of language learners. *Analysis of Intervention in Developmental Disabilities, 3*, 279-293.

Goldstein, H., & Mousetis, L. (1989). Generalized language learning by children with severe mental retardation: Effects of peers' expressive modeling. *Journal of Applied Behavior Analysis, 22*, 245-259.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005) The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179. doi: 10.1177/001440290507100203

Hughes, C., & Rusch, F. R. (1989). Teaching supported employers with severe mental retardation to solve problems. *Journal of Applied Behavior Analysis, 22*, 365-372.

MacManus, C., MacDonald, R., & Ahearn, W. H. (2015). Teaching and generalizing pretend play in children with autism using video modeling and matrix training. *Behavioral Interventions, 30*(3), 191-218. doi: 10.1002/bin.1406

Mahon, C., Lyddy, F., & Barnes-Holmes, D. (2010). Recombinative generalization of subword units using matching to sample. *Journal of Applied Behavior Analysis, 43*(2), 303-307. doi: 10.1901/jaba.2010.43-303

Martella, R. C., Agran, M., & Marchand-Martella, N. E. (1992). Problem solving to prevent work injuries in supported employment. *Journal of Applied Behavior Analysis, 25*, 637-645.

Park, H. S., & Gaylord-Ross, R. (1989). A problem-solving approach to social skills in employment settings with mentally retarded youth. *Journal of Applied Behavior Analysis, 22*, 373-380.

Perez, W. F., & de Rose, J. C. (2010). Recombinative generalization: An exploratory study in musical reading. *The Analysis of Verbal Behavior, 26*, 51-55.

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

Tanji, T., & Noro, F. (2011). Matrix training for generative spelling in children with autism spectrum disorder. *Behavioral Intervention, 26*, 326-339. doi: 10.1002/bin.340

Research and Practice in the Schools
2018, Vol. 5, No. 1, pp. 9-18

ISSN: 2329-5783

Article

Masculine and Feminine Conformity in Lesbian, Gay, Bisexual, and Heterosexual Youth

G. Thomas Schanding, Jr. and Kim A. Case

University of Houston - Clear Lake

Conformity to feminine and masculine norms was examined in a group of heterosexual, gay/lesbian, and bisexual adolescents. Adolescents completed the Conformity to Feminine Norms Inventory or Conformity to Masculine Norms Inventory. Overall, females conformed to a majority of feminine norms, with less conformity reported in regard to investment in appearance for lesbian youth and involvement with children for lesbian and bisexual youth. For gay and bisexual males, less conformity occurred in the areas of disdain for homosexuality, power over women, and violence. Overall, sexual minority youth reported conformity to a majority of gender norms. Implications for practitioners and future research are discussed.

Key words: Adolescents, gender conformity, lesbian, gay, bisexual

Masculine and Feminine Conformity in Lesbian, Gay, and Bisexual Youth

Many disciplines such as psychology, health, and gender studies have supported that males' and females' gendered behavior is influenced by social environments (Addis, Mansfield, & Syzdez, 2010; Bem, 1981). Societal assumptions tend to stereotype gender expression as equal to, and predictive of, gender identity and even sexual orientation. However, that may not fit many youth and their own sense of identity (Lehavot & Lambert, 2007; Saltzberg & Davis, 2010). Of more concern recently are the negative outcomes experienced by some youth who identify as gender nonconforming, especially in schools. Lesbian, gay, bisexual, and transgender (LGBT) youth may experience discrimination, harassment, and even assault due to their sexual orientation (both perceived and real), and/or gender expression, and/or gender identity. This can lead to greater risk for the development of posttraumatic stress disorder; increased suicidality; greater levels of depression and externalizing problems; and generally lower well-being (D'Augelli, Grossman, & Starks, 2006; Friedman, Koeske, Silvestre, Korr, & Sites, 2006; Friedman, Marshal, Stall, Cheong, & Wright, 2008; Williams, Connolly, Pepler, & Craig, 2005).

Unfortunately, social science research also tends to conflate sexual orientation with gender identity and gender expression (Ivory, 2005; Poynter & Washington, 2005). Sexual orientation refers to sex of one's erotic/love/affectionate partner, with some common orientation identities being heterosexual, homosexual, and bisexual (Diamond, 2002). Gender identity is recognition of the perceived social gender attributed to a person – one's self-perception of their gender (Diamond, 2002). Gender expression has been defined as the appearance, gender roles, behaviors, and emotional expression of an individual, with many individuals assuming a label as androgynous, feminine, or masculine (Eagly, 1987; Hoffman, 2001). As children begin to exhibit gender variance from the normative expectations for their assigned sex at birth, these cultural assumptions at times result in assigning minority sexual identities (D'Augelli, Pilkington, & Hershberger, 2002; Gagne & Tewsbury, 1998; Saltzberg & Davis, 2010) rather than a complex analysis of sexual identity, gender identity, and gender expression.

The assumption within the broader cultural context that lesbian, gay, and bisexual (LGB) youth also violate norms via their gender expression contributes to this erroneous conflation of gender identity and sexuality. For example, a prototypically feminine young woman may be assumed heterosexual while her friend with short hair and clothing from the men's section of Old Navy may be assumed to be lesbian. Both assumptions grow directly from gender performance and the perceived levels of each woman's location on the continuum of conformity to nonconformity within prescribed gender systems. The current study aimed to examine these assumptions of LGB gender nonconformity by assessing gender characteristics for LGB and heterosexual youth and comparing their gender expression to national norms.

Gender Expression among LGB Youth

Given the gender performance and conformity expectations imposed culturally and socially, individuals perceived to violate gender systems face not only pressure to conform (Alfieri, Ruble, & Higgins, 1996; Vanderburgh & Forshée, 2003), but also experience marginalization, ridicule, teasing, harassment, abuse, judgment, stigma, and rejection (Burgess, 1999; Chutter, 2007; Horn, 2007; Kimmel & Mahler, 2003; Lobel, 1994; Lobel, Gewirtz, Pras, Shoeshine-Rokach, & Ginton, 1999). As Chutter (2007) noted, these pressures and stigmatizing conditions may lead to negative self-judgments. Some researchers suggested that victimization directed at LGB youth results from gender nonconformity rather than sexual orientation (Horn, 2007; Russel, 2003). D'Augelli and colleagues (2006) examined the impact of gender atypicality on mental health. Gender nonconformity (or atypicality) before the age of 13 (e.g., being called "sissy" or "tomboy," parental discouragement of cross-gender behaviors) predicted higher rates of victimization. In addition, these events before age 13 predicted greater mental health symptomology years later (between ages 15-21).

A few studies explored the distinctions between gender identity, gender conformity/nonconformity, and sexual orientation among adolescents. For example, Horn (2007) presented high school students vignettes describing teenagers with varying levels of gender conformity related to appearance and activities. Horn's study found heterosexual boys exhibited less acceptance of a heterosexual peer for participating in an activity that violates gender norms such as ballet. They rated their acceptance of a heterosexual peer even lower if his appearance failed to conform to gender norms. Toomey, Ryan, Diaz, Card, and Russell (2010) examined adolescent gender nonconformity, victimization in school settings, and psychosocial adjustment among LGBT adults ages 21-25. Along with their reflections on gender nonconformity during school, they also indicated levels of life satisfaction and depression in adulthood. Toomey and colleagues found LGBT status (actual or perceived by peers) mediated the relationship between gender nonconformity and psychosocial adjustment. Given that perceived LGBT status may be derived from gender nonconformity, gender expression may still play a role in the LGBT status variables mediator as analyzed by Toomey et al. (2010) and thus represent another example of conflation of sexual and gender identity.

Examining Assumptions of LGB Gender Nonconformity

Various studies have examined aspects of gender conformity/nonconformity for LGB individuals. Bailey and Zucker (1995) found that gay and lesbian individuals recalled more cross-sex-typed behavior (e.g., toy/activities, cross-dressing, athletics, careers/role models, social reputation for cross-gender behavior) than heterosexual individuals. Additionally, gender nonconforming behaviors, feelings, and interests were also documented for homosexual adults through childhood home videos (Rieger, Linsenmeier, Gygax, & Baily, 2008). While there are limited studies regarding the conformity (or nonconformity) of LGB individuals regarding gender expression, further empirical investigation of gender conformity and nonconformity among LGB youth is needed to determine those individuals' perceptions of conformance to gender norms and how these may or may not align with the perception of others (e.g., school personnel).

The purpose of the current study is to examine LGB and heterosexual school-aged youths' (age 14-21) current feminine and masculine gender characteristics and conformity. Two measures

validated among young adults, the Conformity to Feminine Norms Inventory (CFNI; Mahalik, Morray, Coonerty-Femiano, Ludlow, Slattery, & Smiler, 2005) and the Conformity to Masculine Norms Inventory (CMNI; Mahalik, Locke, Ludlow, Diemer, Scott, Gottfied, & Freitas, 2003), provided a baseline for comparison to the gender characteristics of the general population based on an overall score and subscale scores for particular characteristics. This comparison to a national sample allows for testing of the common assumption of gender nonconformity among LGB and heterosexual youth and young adults.

Method

Participants

Adolescent participants were recruited from a convenience sample of community organizations that provide support to LGBT youth. Organizations serving LGBT youth were emailed by the principal investigator and invited to participate in the project. In total, 33 organizations from 20 states agreed to make the survey available to youth. Ultimately, participants were recruited from 12 states, with a majority of students from Oklahoma (31%), Indiana (16%), and Connecticut (14%). Survey materials were available in paper format in English. This project was reviewed and approved by the principal investigator's university's Institutional Review Board. Youth were informed of their right not to participate and right to withdraw at any time. In order to protect participants from possible breaches of confidentiality, all surveys were completed anonymously upon approval via each agency's staff representative. No names were collected to ensure 100% anonymity. Those youth under the age of 18 provided assent, and agency staff were allowed to provide consent for the youth's participation in the study in the interest of protecting the youth's sexual orientation and/or gender identity/expression.

In all, 144 youth aged 14-21 participated in the project. Those individuals who identified as transgender (n = 8) were excluded from the final analysis of gender conformity as the focus of the study was examination of gender conformity within heterosexual and LGB youth. Ultimately, 100 youth completed the measures fully to be included in the study. The mean age of the sample was 16.88 years (SD = 1.49). Overall, 55% of youth identified their biological sex as female, and 45% identified as male. In terms of ethnicity, the majority of participants identified as Caucasian (58%), 12% as Hispanic, 7% as Native American/Native Alaskan, 2% as Asian/Pacific Islander, 1% as African American, and 18% as Other. Two individuals did not provide their ethnicity. Participants self-identified their sexual orientation: 18% heterosexual, 49% gay/lesbian, and 33% bisexual/other. All youth included in the data analysis identified as cisgender (i.e., not transgender). Data regarding sexual orientation, sex, and gender were collected in multiple-choice format, though youth were also given the option to self-identify as well.

Measures: Conformity to Masculine and Feminine Norms.

Two scales were utilized to measure conformity to gender norms: the Conformity to Feminine Norms Inventory (CFNI; Mahalik et al., 2005) and the Conformity to Masculine Norms Inventory (CMNI; Mahalik et al., 2003). The CFNI contains 84 items, while the CMNI contains 94 items. Items were rated on a four-point Likert scale (*strongly disagree, disagree, agree, strongly agree*). The average of the norm sample for the CFNI was 18-19 years of age (Mahalik et al., 2005); the average age of the norm sample for the CMNI was approximately 20 years old (Mahalik et al., 2003).

These scales provided an overall broad score of Total Conformity to masculine or feminine norms (yielding a *z*-score), but also recognized various subscales of masculinity and femininity that may provide a deeper understanding of the variety of behaviors, feelings, and thoughts (raw scores for each subscale). From the CFNI, eight subscales are provided: 1) Nice in Relationships, 2) Thinness, 3) Modesty, 4) Domestic, 5) Care for Children, 6) Romantic Relationship, 7) Sexual Fidelity, and 8) Invest in Appearance. Internal consistency (coefficient alphas based on the standardization samples) for the CMNI's and CFNI's Total Conformity score are .94 and .88, respectively. Within the CMNI, the following 11 subscales are provided: 1) Winning, 2) Emotional Control, 3) Risk-Taking, 4) Violence, 5) Power Over Women, 6) Dominance, 7) Playboy, 8) Self-Reliance, 9) Primacy of Work, 10) Disdain for Homosexuals,

and 11) Pursuit of Status. The CFNI and CMNI were initially validated with a primarily heterosexual sample (97% for the CFNI [Mahalik et al., 2005]; 96% for the CMNI [Mahalik et al., 2003]). Internal consistencies ranged from .77 to .92 on the CFNI and from .72 to .91 for the subscales on the CMNI.

Measures were completed individually by youth in paper and pencil format. Staff leaders at the various agencies serving LGB youth announced the availability of the study at their site.

Results

Overview of Analysis

Initially, means and standard deviations of scores from the CMNI and CFNI were calculated. For females who completed the CFNI, scores were calculated for those who identified as heterosexual (n = 11), lesbian (n = 22), and bisexual/other (n = 22). For males who completed the CMNI, scores were calculated in the same manner: heterosexual (n = 7), gay (n = 27), and bisexual/other (n = 11).

T-scores were calculated based on each subgroup's mean compared to the norms of the respective instrument's standardization sample (see Tables 1 & 2). Initially, differences were calculated for the CFNI and CMNI Total score to determine if there were differences based on a broad measure of gender conformity. Ultimately, *t*-tests were also conducted for each subscale, as the subscales provide a richer context of the thoughts, feelings, and behaviors of the participating youth. A Bonferroni correction was applied to account for the multiple comparisons made among the scales to determine meaningful differences. Differences were considered significant at the level of $p < .00185$ or less for comparisons with the CFNI measure and $p < .00138$ for comparisons with the CMNI measure.

Findings

Females. A *t*-test was performed comparing the means of the Total CFNI score for each of the three groups (e.g., heterosexual, lesbian, and bisexual/other) to the standardization sample. When examining the Total CFNI score, females identifying as heterosexual indicated less conformity to feminine norms (M = -.80) than the standardization sample, (M = 0), t (742) = 2.59, p = .0097, d = .62. Female youth in the current study identifying as heterosexual also differed significantly from the standardization sample in two subscales. Heterosexual females indicated lower scores on the subscale Involvement with Children (M = 19.82) than the standardization sample (M = 26.51), t (742) = 3.54, p = .0004, d = 1.15, and on the Domestic subscale, (M = 10.64 v. M = 14.64), t (742) = 3.63, p = .0003, d = .83. No other differences were noted in subscales between the heterosexual youth in the current sample and the standardization sample.

Table 1
Overview of Mean (SD) CFNI Scores for Females Identifying as Heterosexual, Lesbian, and Bisexual/Other

	Heterosexual (n=11)	Lesbian (n=22)	Bisexual/ Other (n=22)
CFNI Total Score	-.80 (1.54)	-.31 (.96)	-.38 (.90)
Having Nice Relationships	38.36 (3.80)	40.16 (6.11)	37.86 (5.10)
Involvement with Children	19.82*** L (5.36)	19.41**** L (8.47)	22.07*** L (4.99)
Thinness	19.36 (7.47)	19.47 (7.95)	20.70 (6.67)
Sexual Fidelity	18.00 (2.24)	21.31 (4.92)	19.05 (5.53)
Modesty	12.82 (3.68)	13.09 (4.60)	14.59 (2.75)
Involvement in Romantic Relationships	13.91 (3.08)	17.28 (4.25)	14.00 (4.87)
Domestic	10.64*** L (5.03)	15.67 (4.97)	14.95 (5.13)
Investment in Appearance	12.18 (4.31)	8.75*** L (3.53)	11.66 (5.14)

Note: CFNI Total Score is a *z* score. Subscales are raw scores. H = Higher than CFNI norm group; L = Lower than CFNI norm group. * $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$

Table 2
Overview of Mean (SD) CMNI Scores for Males Identifying as Heterosexual, Gay, and Bisexual/Other

	Heterosexual (*n*=7)	Gay (*n*=27)	Bisexual/ Other (*n*=11)
CMNI Total Score	-.46 (.89)	-1.10**** L (1.19)	-1.18**** L (1.35)
Winning	14.43 (5.77)	13.85 (5.45)	13.73 (7.77)
Emotional Control	15.29 (3.64)	13.22 (9.88)	12.82 (6.90)
Risk Taking	17.14 (2.55)	16.36 (6.19)	14.27 (4.54)
Violence	12.86 (5.18)	10.33 (6.34)	7.00**** L (6.16)
Power Over Women	6.86 (4.53)	5.70**** L (4.97)	4.82**** L (3.97)
Dominance	5.29 (1.80)	5.37 (2.39)	5.82 (3.22)
Playboy	10.71 (3.55)	8.68 (6.61)	10.00 (7.84)
Self-Reliance	6.43 (3.41)	7.89*H (4.58)	5.91 (4.44)
Primacy of Work	10.71 (2.87)	10.13 (4.32)	9.18 (4.45)
Disdain for Homosexuality	13.14 (4.01)	4.33**** L (4.80)	10.64*** L (7.39)
Pursuit of Status	10.57 (.54)	11.47 (2.58)	11.09 (3.86)

Note: CMNI Total Score is a *z* score. Subscales are raw scores. H= Higher than CMNI norm group; L = Lower than CMNI norm group. * $p < .05$; ** $p < .01$; *** $p < .001$; **** $p < .0001$

No significant differences were found for females identifying lesbian and bisexual/other with regard to the Total CFNI score. However, investigation of the subscales revealed more specific differences with regard to feminine norms. Females identifying as lesbian received lower scores on the Involvement with Children subscale (M = 19.41) than the standardization sample (M = 26.51), t (750) = 4.85, p = .0001, d = .95. Females identifying as lesbian also received lower scores on the Investment in Appearance subscale (M = 8.75) compared to the standardization sample (M = 12.01), t (749) = 3.65, p = .0003, d = .89. Female youth identifying as bisexual/other received lower scores on the subscale Involvement with Children (M = 22.07) compared to the standardization sample (M = 26.51), t (753) = 3.31, p = .001, d = .78. All other subscales comparing youth identifying as lesbian or bisexual/other to the CFNI standardization sample were nonsignificant.

Males. A *t*-test was performed comparing the means of the Total CMNI score for each of the three groups (e.g., heterosexual, gay, and bisexual/other) to the standardization sample. Youth identifying as heterosexual had no significant differences in the Total CMNI score or subscale scores when compared to the standardization sample of the CMNI. Youth identifying as gay received a lower Total CMNI score (M = -1.10) compared to the standardization sample (M = 0), t (777) = 5.59, p = .0001, d = 1.05. Additionally, youth identifying as bisexual/other received a lower total CMNI score (M = -1.18) compared to the standardization sample, t (761) = 3.88, p = .0001, d = .99.

Males identifying as gay also reported less adherence to certain areas of masculinity as measured by the CMNI. Youth identifying as gay received a lower score on the Winning subscale (M = 13.85) compared to the standardization sample (M = 16.91), t (777) = 3.06, p = .0023, d = .58. Additionally, gay males reported lower scores compared to the standardization samples on the subscales Power Over Women (M = 5.7 v. M = 10.59; t (777) = 5.58, p = .0001, d = 1.04), Playboy (M = 8.68 v. M = 12.06; t (777) = 5.58, p = .0046, d = .53), and Disdain for Homosexuality (M = 4.33 v. M = 17.74; t (777) = 10.38, p = .0001, d = 2.31).

In examining males identifying as bisexual or other, these youth received lower scores related to Violence (M = 7.00) compared to the national sample (M = 12.38), t (761) = 4.43, p = .0001, d = 1.04. Lower scores for bisexual/other youth were also reported with regard to Power Over Women (M = 4.82 v. M = 10.59), t (761) = 4.27, p = .0001, d = 1.36, as well as Disdain for Homosexuality (M = 10.64 v. M = 17.74), t (761) = 3.51, p = .0005, d = 1.01.

Discussion

The purpose of this study was to examine LGB and heterosexual school-aged youths' perceptions of their own conformity (or nonconformity) to masculine and feminine gender norms as compared to a normative sample. Overall, LGB males and females reported similar behaviors, feelings, and thoughts as a majority heterosexual sample regarding conformity to masculine and feminine norms. Previous investigations have emphasized some of the departures of LGB youth from stereotypical masculine and feminine norms (i.e., Bailey & Zucker, 1995; Rieger et al., 2008); however, research has not fully captured the extent to which LGB youth may embody masculinity and femininity.

Female LGB and Heterosexual Youth

For female participants, only three gender subscales differed from the national sample. All female youth (lesbian, bisexual, and heterosexual) in the current sample reported less conformity than the national norm sample for the subscale Involvement with Children (e.g., I would baby-sit for fun). However, the mean difference was more pronounced for females identifying as lesbian than those identifying as heterosexual or bisexual/other. The current sample's somewhat younger age in comparison to the standardization sample may explain the lower levels of involvement with children reported here. In general, females identifying as lesbian indicated less conformity than the national norm sample in attitudes, behaviors, and thoughts regarding the Investment in Appearance subscale (e.g., I never wear makeup (reverse); It is important to look physically attractive in public). Female youth identifying as heterosexual indicated less conformity than the national norm sample with regard to the Domestic subscale (e.g., I enjoy spending time making my living space look nice; It is important to keep your living space clean).

Similarities to the national sample appeared more often than differences among sexual minority female youth for gender subscale scores. Females identifying as lesbian or bisexual/other reported conforming for a majority of subscales on the CFNI. This finding might seem unexpected and run counter to typical cultural assumptions about gender expression among lesbian and bisexual women. Lesbian participants' self-reports were similar to the national norm sample on the following CFNI subscales: Having Nice Relationships, Thinness, Sexual Fidelity, Modesty, Involvement in Romantic Relationships, Domestic. Bisexual female participants' responses were similar to established national norms on the following subscales: Having Nice Relationships, Thinness, Sexual Fidelity, Modesty, Involvement in Romantic Relationships, Domestic, and Investment in Appearance.

Male LGB and Heterosexual Youth

Overall, male youth identifying as gay or bisexual/other were less conforming to masculine gender norms than male youth identifying as heterosexual. Across male participants, several gender subscales differed in comparison to the national sample. Males identifying as gay reported less conformity with masculine norms, compared to the national norm sample, with regard to the following scales: Winning (e.g., I will do anything to win; Winning isn't everything, it's the only thing; The best feeling in the world comes from winning), Power Over Women (e.g., Women should be subservient to men; I will only be satisfied when women are equal to men [reverse]); Playboy (e.g., Emotional involvement should be avoided when having sex; Long term relationships are better than casual sex [reverse]); and Disdain for Homosexuality (e.g., I try to avoid being perceived as gay; I would be furious if someone thought I was gay). Males identifying as bisexual/other in the current sample reported less conformity than the national norm sample with regard to the following subscales: Violence (e.g., I am willing to get into a physical fight if necessary; No matter what the situation, I would never act violently [reverse]); Power Over Women; and Disdain for Homosexuality.

In comparison to the national sample, gay and bisexual male youth responded similarly on several masculinity subscales. Gay male participants' self-reports were similar to the national norm sample on the following CMNI subscales: Emotional Control, Risk Taking, Violence, Dominance, Self-Reliance, Primacy of Work, Pursuit of Status. Bisexual male participants' responses were similar to established

national norms on the following subscales: Winning, Emotional Control, Risk Taking, Dominance, Playboy, Self-Reliance, Primacy of Work, Pursuit of Status.

Males identifying as heterosexual did not differ significantly from the comparison national sample. Given that the sample was recruited via organizations that support LGB youth, heterosexual male reports of conformity to Disdain for Homosexuality were somewhat surprising. Social desirability concerns might predict that heterosexual male youth involved in LGB organizations would avoid anti-gay type responses. However, this particular social context might also trigger fears of being labeled as gay and therefore lead to higher rates of masculine conformity to anti-gay responses.

Limitations and Future Directions

The small sample size for each individual category (e.g., 11 bisexual/other males) represents one limitation for the CFNI and CMNI comparisons studied above, preventing generalizations to all LGB and heterosexual youth; however, it should be noted that it is sometimes difficult to recruit LGB youth for studies given concerns regarding issues of confidentiality, consent, and access in general to this population.

Current participants completed gender subscales for their culturally assigned gender. In other words, females did not respond to masculinity subscales and male participants did not complete the femininity subscales. Future studies might present both feminine and masculine measures to all participants for a broader comparison across the gender spectrum. Considering the findings of many gender similarities among LGB participants and the national sample, future studies focusing on youth may contribute to challenging assumptions of gender nonconformity among sexual minority youth. It should be noted that the measures used in the current study do adhere to a feminine-masculine gender binary, which may not adequately describe all youth who participated in the current study.

The heterosexual youth involved in LGB organizations represent the under-studied "allies" of the sample. Additional studies on heterosexual youth acting as allies to their LGB peers may develop understanding of the connections among ally attitudes, behaviors, and gender conformity or nonconformity. The current finding that heterosexual males acting as allies still reported disdain for homosexuality needs further exploration as well.

It should be noted that the majority of lesbian youth surveyed in this study reported an identity consistent with a conformity view of femininity. The CFNI Total score for adolescent lesbians was similar to that of the standardization sample of the CFNI. There were no differences related to these youths' perceptions of their relationships with others, thinness, sexual fidelity, involvement in romantic relationships, or domesticity. While there may be a stereotype from society that lesbians may be more gender nonconforming, this was not found in the current sample. Similar to the findings for lesbian youth, gay male adolescents overall reported conforming to a majority of masculine norms more than what may be perceived by society at large. Within this sample, gay males reported conformity on nine out of 11 dimensions of masculinity. Longitudinal research with LGB youth exploring gender conformity and sexual orientation from adolescence to early adulthood/adulthood is warranted to examine any developmental changes associated with gender conformity and sexual orientation. Perhaps as adolescents mature, they become more likely to explore and endorse gender nonconformance.

The gender identities of these youth are likely more complex than common stereotypes suggest and deserve further investigation. This type of research is necessary for understanding the intricate link between sexual orientation and gender expression – especially as it may pertain to LGB youths' experience of victimization within schools and communities. These findings would also be useful to educators, practitioners, and clinicians providing services to LGB youth. If school personnel assume an adolescent is heterosexual or homosexual based on gender conformity/ nonconformity, they may unwittingly enact microaggressions on clients by asking questions that imply a certain sexual orientation.

As suggested by Horn (2007) and Russel (2003), harassment and victimization directed at LGB youth may result from gender nonconformity rather than sexual orientation; however, not all nonconformance may be equal when considering

the various dimensions of femininity and masculinity. Further research may wish to explore (through direct measurement or possibly through experimental manipulation of gender conformity vignettes) attitudes and behaviors of youth of varying levels of conformity/nonconformity on specific domains. Perhaps a female who violates the feminine dimension of Investment in Appearance (expressing her gender more as male) would be more likely to be victimized than if she exhibited less conformity to Domestic traits. This may also be hypothesized perhaps if LGB youth were to be more conforming to characteristics of the opposite sex – such as a lesbian female who is perceived as having masculine characteristics of Dominance and Playboy (from the CMNI) or a gay male who is perceived as having feminine characteristics of Investment in Appearance and Having Nice Relationships (from the CFNI). In other words, nonconformity to certain stereotypical masculine or feminine characteristics may lead to harsher victimization. Further research would provide support needed to explore this hypothesis.

Implications for Educators

As noted before, LGB youth are often at risk for a multitude of negative mental health outcomes (e.g., D'Augelli et al., 2006; Friedman et al., 2006; Friedman et al., 2008; Williams et al., 2005). As affirmed by the American Psychological Association (APA) and the National Association of School Psychologists (NASP), all youth, including those who are sexual or gender minority children and adolescents, have the right to equal opportunity and a safe environment within all public educational institutions (APA & NASP, 2015). We recommend that school personnel receive training regarding socialization of gender roles and the importance of addressing gender bias when working with students. We also recommend incorporating these complex distinctions among sexuality and gender expression within graduate training programs for school psychologists, teachers, and administrators in education. School personnel would benefit from further considering their own expectations of behavior related to gender conformity and sexual orientation. School psychologists should be aware of and address stereotyped beliefs held by educators (and their own beliefs) when discussing child and adolescent development. Educators who do not provide a safe, supportive environment, or continually rely on detrimental stereotypes, miss the opportunity to recognize the commonalities of sexual and/or gender minority youth with their peers and celebrate diverse viewpoints. The stigma associated with these negative beliefs contributes to poorer mental health outcomes for LGB youth (Meyer, 2003; Meyer, Schwartz, & Frost, 2008).

Conclusions

While there may be hypotheses about LGB receiving harassment, discrimination, and bullying by not conforming to gender norms, these data indicate that by and large LGB youth self-report to conform to a majority of gender-conforming behaviors. Adolescence is a major time of identity development, where sex and gender roles are developing. However, these roles may be fluid to some extent over time and are continually refined (Deutsch, 2007; West & Zimmerman, 1987). As always, professionals consulting with families and providing direct services to adolescents (including school personnel, health service providers, etc.) should carefully examine their own preconceptions of sexual orientation and gender conformity/nonconformity as there is a need to provide affirming practices for gender nonconforming youth when considering health and mental healthcare (Case & Meier, 2014; Edwards-Leeper, Leibowitz, & Sangganjanavanich, 2016; Hidalgo et al., 2013). Educators, practitioners, and clinicians are uniquely situated to behave as adult allies to gender creative and gender expansive identities and expressions among LGB youth. A more nuanced understanding of sexual orientation and gender nonconformity should lead to more sensitivity to the needs of LGB youth and increased effectiveness of services.

References

Addis, M. E., Mansfield, A. K., & Syzdek, M. R. (2010). Is "masculinity" a problem?: Framing the effects of gendered social learning in men. *Psychology of Men & Masculinity, 11*, 77–90. 10.1037/a0018602

Alfieri, T., Ruble, D. N., & Higgins, E. T. (1996). Gender stereotypes during adolescence: Developmental changes and the transition to junior high school. *Developmental Psychology, 32*(6), 1129–1137.

American Psychological Association & National Association of School Psychologists. (2015). *Resolution on gender and sexual orientation diversity in children and adolescents in schools*. Retrieved from http://www.apa.org/about/policy/orientation-diversity.aspx

Baily, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology, 31*(1), 43-55.

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*, 354–364. doi:10.1037/0033-295X.88.4.354

Burgess, C. (1999). Internal and external stress factors associated with the identity development of transgendered youth. *Social Services with Transgendered Youth*, *10*(3/4), 35–47.

Case, K., & Meier, S. C. (2014). Developing allies to transitioning and gender-nonconforming youth: Training for counselors and educators. *Journal of LGBT Youth, 11*, 62-82.

Chutter, K. (2007). Opening our awareness to heterosexist and homophobic attitudes in society. *Relational Child and Youth Care Practice*, *20*(3), 22–27.

D'Augelli, A. R., Grossman, A. H., & Starks, M. T. (2006). Childhood gender atypicality, victimization, and PTSD among lesbian, gay, and bisexual youth. *Journal of Interpersonal Violence*, *21*(11), 1-21.

D'Augelli, A. R., Pilkington, N. W., & Hershberger, S. L. (2002). Incidence and mental health impact of sexual orientation victimization of lesbian, gay, and bisexual youths in high school. *School Psychology Quarterly, 17*, 148-167. doi:10.1521/scpq.17.2.148.20854

Deutsch, F. M. (2007). Undoing gender. *Gender & Society*, *21*(1), 106-127. doi:10.1177/0891243206293577

Diamond, M. (2002). Sex and gender are different: Sexual identity and gender identity are different. *Clinical Child Psychology and Psychiatry, 7*(3), 1359-1045.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

Edwards-Leeper, L., Leibowitz, S., & Sangganjanavanich, V. F. (2016). Affirmative practice with transgender and gender nonconforming youth: Expanding the model. *Psychology of Sexual Orientation and Gender Diversity, 3(2)*, 165-172. doi:10.1037/sgd0000167

Friedman, M. S., Koeske, G. F., Silvestre, A. J., Korr, W. S., & Sites, E. W. (2006). The impact of gender-role nonconforming behavior, bullying, and social support on suicidality among gay male youth. *Journal of Adolescent Health, 38*, 621-623.

Friedman, M. S., Marshal, M. P., Stall, R., Cheong, J., & Wright, E. R. (2008). Gay-related development, early abuse, and adult health outcomes among gay males. *AIDS and Behavior, 12*, 891-902.

Gagne, P., & Tewksbury, R. (1998). Rethinking binary conceptions and social constructions: Transgender experiences of gender and sexuality. In M. T. Segal & D. Vasilikie (Eds.), *Advances in Gender Research* (Vol. 3; pp. 73–102). Greenwich, CT: JAI Press.

Hidalgo, M. A., Ehrensaft, D., Tishelman, A. C., Clark, L. F., Garofalo, R., Rosenthal, S. M., Olson, J. (2013). The gender affirmative model: What we know and what we aim to learn. *Human Development, 56*, 285-290. doi:10.1159/000355235

Hoffman, R. M. (2001). The measurement of masculinity and femininity: Historical perspectives and implications for counseling. *Journal of Counseling & Development, 79*, 472–485. doi:10.1002/j.1556-6676.2001.tb01995.x

Horn, S. S. (2007). Adolescents' acceptance of same-sex peers based on sexual orientation and gender expression. *Journal of Youth and Adolescence*, *36*, 363-371.

Ivory, B. (2005). LGBT students in community college: Characteristics, challenges, and recommendations. *New Directions for Student Services, 111*, 61-69. doi: 10.1002/ss.174

Kimmel, M. S., & Mahler, M. (2003). Adolescent masculinity, homophobia, and violence: Random school shootings, 1982–2001. *American Behavioral Scientist*, *46*, 1439–1458.

Lehavot, K., & Lambert, A. J. (2007). Toward a greater understanding of antigay prejudice: On the role of sexual orientation and gender role violation. *Basic and Applied Social Psychology, 29*, 279–292. doi: 10.1080/01973530701503390

Lobel, T. (1994). Sex typing and the social perception of gender stereotypic and nonstereotypic behavior: The uniqueness of feminine males. *Journal of Personality and Social Psychology*, *66*(2), 379–385.

Lobel, T., Gewirtz, J., Pras, R., Shoeshine-Rokach, M., & Ginton, R. (1999). Preadolescents' social judgments: The relationship between self-endorsement of traits and gender-related judgments of female peers. *Sex Roles*, *40*(5), 483-498.

Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P .J., Gottfried, M., & Freitas, G. (2003). Development of the Conformity to Masculine Norms Inventory. *Psychology of Men & Masculinity*, *4*(1), 3-25. doi:10.1037/1524-9220.4.1.3

Mahalik, J. R., Morray, E. B. Coonerty-Femiano, A., Ludlow, L. H., Slattery, S. M., & Smiler, A. (2005). Development of the Conformity to Feminine Norms Inventory. *Sex Roles*, *52*(7/8), 417-435. doi:10.1007/s11199-005-3709-7

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129*, 674-679. doi:10.1037/0033-2909.129.5.674

Meyer, I. H., Schwartz, S., & Frost, D. M. (2008). Social patterning of stress and coping: Does disadvantaged social statuses confer more stress and fewer coping resources? *Social Science & Medicine, 67*, 368-379.

Poynter, K. J., & Washington, J. (2005). Multiple identities: Creating community on campus for LGBT students. *New Direction for Student Services, 111*, 41-47. doi: 10.1002/ss.172

Rieger, G., Linsenmeier, J. A. W., & Gygax, J. M. (2008). Sexual orientation and childhood gender nonconformity: Evidence from home videos. *Developmental Psychology, 44*(1), 46-58. doi:10/1037/0012-1649.44.1.46

Russell, S. T., (2003). Sexual minority youth and suicide risk. *American Behavioral Scientist, 46*(9), 1241–1257.

Saltzburg, S., & Davis, T. S. (2010). Co-authoring gender-queer youth identities: Discursive tellings and retellings. *Journal of Ethnic & Cultural Diversity in Social Work, 19*(2), 87-108.

Toomey, R. B., Ryan, C., Diaz, R. M., Card, N. A., & Russell, S. T. (2010). Gender-nonconforming lesbian, gay, bisexual, and transgender youth: School victimization and young adult psychosocial adjustment. *Developmental Psychology, 46*(6), 1580-1589.

Vanderburgh, R., & Forshée, A. S. (2003, November). *Depathologizing transgender youth identity: The identity project*. Paper presented at the American Public Health Association 131st Meeting on Promoting a Health Future: LGBT Youth Research & Programming, San Francisco, CA.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender and Society, 1*, 125-151.

Williams, T., Connolly, J., Pepler, D., & Craig, W. (2005). Peer victimization, social support, and psychosocial adjustment of sexual minority adolescents. *Journal of Youth and Adolescence, 34*, 471-482.

Research and Practice in the Schools
2018, Vol. 5, No. 1, pp. 19-28

ISSN: 2329-5783

Article

Ready or Not Here They Come: An Overview of National School-Readiness Screening Policy

Elena Lilles Diamond
Lewis & Clark College

Shelley R. Hart
California State University, Chico

Jenna Carlson
Lewis & Clark College

Alex Roth
Lewis & Clark College

In response to increasing national prioritization of school-readiness (most recently Race to the Top funding from U.S. Department of Education), states have begun their own school-readiness initiatives to ensure that all students enter kindergarten ready to learn. This brief report provides a snapshot of state-level school readiness practices sorted into three categories: (1) *active school readiness screening*; (2) *partial school readiness screening*; and (3) *no school readiness screening*. Results indicated that 51% of the states (n = 26) had specific school readiness screening procedures, 33% (n = 17) fell into the partial school readiness screening category, and 16% (n = 8) fell in the no school readiness screening category. Basic demographic information for the states indicated no significant differences in number of districts, number of students enrolled, or annual median income. The importance of school readiness screening emphasizes the need for a clear, national strategy.

Keywords: School Readiness, Screening, Policy

A growing database of early education and school readiness literature illustrates that early achievement and readiness for kindergarten is a significant predictor of later academic success (Graziano, Reavis, Keane, & Calkins, 2007; Lonigan 2006). For example, children who have a broad base of school readiness experiences have been found to acquire complex skills more rapidly than those who do not (Bowman, Donovan, & Burns, 2000). Similarly, research suggests that school entry abilities in math, reading, and attention predict later academic achievement (Duncan et al., 2007), and physical, behavioral, and social-emotional skills are a similarly important foundation for school success (Davies, Janus, Duku, & Gaskin, 2016). Academic performance trajectories have been found to stabilize early in one's academic career (Lilles et al., 2009; Torgesen & Burgess, 1998) and students from disadvantaged backgrounds often fall further behind their peers as they progress through school, further highlighting the importance of school readiness for all students

Author Note: Correspondence for this article should be addressed to: Elena Diamond, Lewis & Clark College, Graduate School of Education and Counseling, 0615 SW Palatine Hill Road, Portland, Oregon 97219. Email: ediamond@lclark.edu

(Rumberger & Arellano, 2009). Students who exhibit low academic achievement are at a greater risk for school failure, and exhibit higher dropout rates (Bridgeland, DiJulio, & Morison, 2006). In addition, the literature supports that prevention and early intervention efforts addressing school-related and social-emotional variables are a more effective use of resources than later remediation (Karoly, Kilburn, & Cannon, 2005; U.S. Department of Education, 2015).

While historically school readiness may not have been within the purview of a school psychologist's typical daily practice (i.e., outside the special education realm), over the last decade there have been calls to expand the service provision more deeply into early education (e.g., Hojnoski & Missall, 2006). Additionally, with the emphasis on prevention, early intervention, and screening (e.g., Response to Intervention [RtI], Multi-Tiered Systems of Support [MTSS]) in the field and in training programs (Canter, 2006; National Association of School Psychologists, 2010), it follows that school psychologists are primed and prepared to engage in school readiness screening.

School Readiness Legislation

In recent years, U.S. policies and practices at the national, state, and local levels have reflected a renewed focus on school readiness, observed in policy reform and funding opportunities. This is highlighted in the first of the National Education Goals, developed in 1990, which sets the intention that "all children in America will enter school ready to learn." Funding opportunities have followed, including the Early Learning Challenge Fund in 2009, which incentivized reforms of state early learning programs. These federal grants were designed to promote components of early learning systems including: (1) systems to facilitate screening and referrals, and (2) systems to promote age and developmentally-appropriate curriculum and assessments used to guide practice, improve programs, and inform kindergarten readiness (U.S. Department of Education, 2009). Another funding source included the 2014 Preschool Development Grants generated to fund the development of high-quality preschool programming and infrastructure aimed to prepare states for Preschool for All initiatives (U.S. Department of Education, 2016).

In addition to federal legislation, many states have significantly expanded funding efforts to support preschool programming. The Education Commission of the United States reports that overall, states have raised funding of preschool programs by 47% in the past five years. Further, 30 states budgeted an increase totaling $480 million to support preschool programs in 2016-17 fiscal year, equating to an overall increase of 6.8% since 2015-16 (Diffey, Parker, & Atchison, 2017). Several states have also implemented legislation that supports Preschool for All or Universal Preschool programming, either now or in the future. For example, Florida is one state that has legislation in place to support Universal Preschool for all 4-year-old children (Florida Office of Early Learning, 2017).

School Readiness Screening

Given the national interest in school readiness, and the ability to identify a child's readiness for kindergarten with the use of screening tools, many states have developed formal policies regarding school readiness screening practices (Daily, Halle, Burkhauser, & Child, 2010). Broadly defined, school readiness screening refers to the use of a brief assessment tool to evaluate children on domains associated with successful transition into kindergarten and later achievement such as cognitive, academic, social-emotional, language, and/or health, to identify students who many need additional interventions, resources, or evaluation prior to, or at the beginning of, kindergarten (Thurlow & Gilman, 1999). School readiness screening is considered a wide-reaching, cost-effective practice that provides a snapshot of a child, or cohort of children, at the time of the screening. There are a variety of tools that can be used to facilitate the screening process including formal norm-referenced standardized measures, informal measures, and naturalistic observations. Some school districts, and even states, have developed their own specific school readiness screeners to gather information deemed most relevant to meet their screening needs. For example, Alaska uses the Alaska Developmental Profile, a state-developed observational tool that gathers information across the domains of (a) physical health and wellbeing, (b) social emotional development, (c) approaches to

learning, (d) cognition and general knowledge, and (e) communication language and literacy, at the beginning of kindergarten or the beginning of first grade for students who were not enrolled in kindergarten (for details visit the Alaska Department of Education and Early Development at https://education.alaska.gov).

School readiness screening data can be used in the short-term to identify immediate needs, as well as aid in ongoing student monitoring. For instance, school readiness screening can be considered an entry point to an RtI or MTSS model, and can be used to facilitate curriculum and intervention planning, the development of individualized instruction, and ongoing progress monitoring (Diamond, Quirk, & Furlong, 2016). Screening can also be used in part of large-scale program evaluation which may inform funding and policy decisions. As school psychologists are trained in data-based decision-making, RtI and MTSS models, and their daily practice likely includes these practices, they are perfectly primed to engage in these discussions and to encourage systems to adopt systematic school readiness activities.

Still relatively new, school readiness screening practices vary dramatically from state to state. Many states have developed state-wide screening practices, and in some instances these practices have translated into formal policy. This manuscript provides a snapshot of the current screening practices and policy at the state level across the United States as reported in spring of 2016 to showcase the growing focus on school readiness at a policy level, as well as to bring focus to the gaps of formal screening practices at a national level. As such, for the current project, school readiness screening is broadly defined as any systematic practice of evaluating student readiness either prior to kindergarten or during the kindergarten year, that includes one or more of the broad factors associated with a successful transition into kindergarten and later academic success as largely defined in the school readiness literature (e.g., cognitive, social-emotional, behavioral, academic).

Methods

The current study sought to evaluate the pulse of the nation regarding current school readiness screening practices, specifically to identify which states have implemented formal policy regarding school readiness screening. Information was gathered for all 50 states and the District of Columbia (DC) regarding current state mandated school readiness screening practices and policy regulations. Data were collected first via a thorough review of state Administrative Rules, state Administrative Codes, or similar documents that hold state bylaws for all 50 states and DC. For states in which no formal policy was identified in this review, a secondary review was conducted by searching state Department of Education (DOE) websites to identify if any state-wide screening practices were being implemented that had not yet been translated into formal policy.

Documents and websites were searched using the key terms “kindergarten assessment,” “kindergarten entrance evaluation,” “kindergarten evaluation,” “kindergarten screening,” “kindergarten readiness,” “preschool assessment,” “preschool readiness,” “preschool evaluation,” and “school readiness,” as these terms are often used interchangeably within school readiness literature, or are associated with school readiness or kindergarten readiness screening or evaluation practices (e.g., Costenbader, Rohrer, & Difonzo, 2000; Daily, et al., 2010; Dockett & Perry, 2002). Documents were evaluated for information documenting the state-wide use of screening or assessment prior to, or at the beginning of kindergarten to evaluate student readiness.

Using available data, states were sorted into one of three categories: (a) *active school readiness screening*; (b) *partial school readiness screening*; and (c) *no formal school readiness screening*. The *active school readiness screening* category is defined to include states in which there is an active state-wide school readiness practice in place, enforced by state policy, and documented in the state’s Administrative Rules or comparable document. States were coded to be included in this category if there was an active Administrative Rule (or similar) stating that the use of a kindergarten/school readiness screener was mandated at the state level. For example, in the state of Oregon, the Oregon Administrative Rule 581-022-2130 states that, “beginning with the 2013-2014 school year, all school districts shall administer the kindergarten assessment to students

who are enrolled in kindergarten." The rule further outlines elements of the screener including domains to be measured, applicability to all students, and alignment to Oregon's Early Learning and Development Standards and Common Core State Standards.

The *partial school readiness* category is defined to include states in which school readiness screening practices are in place or partially in place, but there is not a state mandate outlined in the state's Administrative Rules or comparable documents, or states that have a future-dated policy that is not currently active. States were coded to be included in this category if they met the following criteria: (a) states that have a state-wide school readiness screening practice, however there is not a state regulation mandating this practice (e.g., state-wide kindergarten readiness program is highlighted on the DOE website but there is no formal legal precedent documented in the state Administrative Rules or similar document); (b) states that have a state law mandate to be rolled out at a future time (e.g., all districts will implement kindergarten readiness screening by the 2019-20 academic year); or (c) there is mention of school readiness screening practice; however, the practice is left up to the district to determine if and how the practice will be implemented. For example, according to the Missouri Department of Elementary and Secondary Education, the state of Missouri recommends but does not mandate the use of the Desired Results Developmental Profile (DRDP) as a school readiness screener. Likewise, according to the Pennsylvania Department of Education there are no formal requirements for kindergarten assessment in Pennsylvania state law, however all school districts are invited to register to use the Kindergarten Entry Inventory.

The *no formal school readiness* category is defined to include states in which there are no data available to suggest that a state-wide school readiness screening practice is in place, or that there are any plans to implement such a practice in the near future. States were coded to be included in this category if kindergarten readiness screening is not documented in the state's Administrative Rules or similar documents, or on the state DOE website. For example, a review of the Texas Administrative Codes, the Texas Education Codes as well as the Texas Education Agency website did not yield any formal mention of systematic school readiness screening. This does not mean that the practice of school readiness screening does not exist in the state, but rather, there is no documented formal screening practice at a state level.

As a means to ensure accuracy, a second researcher completed an independent search of the Administrative Codes and Department of Education websites for each state to confirm findings. Inter-rater reliability was evaluated and there was no disagreement between the first and second reviewers for each state.

Finally, the National Education Association's (2016) report on the rankings of the states and school statistics was utilized to attempt to begin describing the states who have formalized screening procedures and those who do not. A one-way between subjects ANOVA was conducted to investigate the differences in annual median household income level, annual enrollment in school, and number of districts, all reported at the state level. These indicators were chosen as they were readily available and accessible data—consistently reported for all 50 states and DC; however, it was also hypothesized that resources, or lack thereof, might interfere with policy directed toward school readiness activities. While each state differs on funding, typically taxes from income are a source of monies that are available for states to allocate to districts (Center for Public Education, n.d.). If states have less money available, initiatives that are not mandated in order to receive federal dollars, such as school readiness activities, may receive less priority (Condron, & Roscigno, 2003; Hedges, Laine, & Greenwald, 1994). Likewise, when there are many districts in the state, funding might be spread thin amongst those districts. Finally, schools that are overcrowded may be simply trying to survive with the mandates that are already in place. States in these situations (i.e., lower annual median income, higher numbers of students enrolled, larger overall districts) might then not focus on recommendations to enact policy related to school readiness screening (Hedges, et al., 1994). This analysis was conducted using SPSS v. 22 (SPSS, 2013).

Table 1

States by School Readiness Screening Category

Active School Readiness Screening (*n*=26)		Partial School Readiness Screening (*n*=17)		No Formal School Readiness Screening (*n*=8)
Alaska	New Mexico	Arizona	New Hampshire	Alabama
Arkansas	New York	California	Pennsylvania	Idaho
Colorado	North Carolina	Connecticut	Rhode Island	Montana
Delaware	North Dakota	Georgia	Virginia	Nebraska
District of Columbia	Ohio	Hawaii		South Dakota
Florida	Oklahoma	Illinois		Tennessee
Iowa	Oregon	Indiana		Texas
Kentucky	South Carolina	Kansas		West Virginia
Louisiana	Utah	Maine		
Maryland	Vermont	Massachusetts		
Minnesota	Washington	Michigan		
Mississippi	Wisconsin	Missouri		
New Jersey	Wyoming	Nevada		

Table 2

Descriptive Information based on Readiness Screening Processes Identified per State

Screening Procedures	Median Annual Household Income Mean (SD)	Enrollment Mean (SD)	Number of Districts Mean (SD)	Total Number of Students
Active (*n*=26)	$57,280.50 ($10,343.01)	804,820.73 (697,717.26)	255.77 (241.85)	20,925,339
Partial (*n*=17)	$59,328.88 ($7,665.33)	1,215,769.76 (1,427,562.52)	380.29 (314.00)	19,350,942
None (*n*=8)	$51,211.75 ($6,081.23)	1,011,182.88 (1,724,075.81)	311.75 (381.64)	8,089,463

Note. SD = Standard Deviation.

Results

Table 1 lists results from the 50 states and DC based on the authors' attempts at finding procedures for screening discussed at the state level. Descriptive analysis indicated that 51% of the states (n = 26) fell into the *active school readiness screening* category, 33% (n = 17) fell into the *partial school readiness screening* category, and 16% (n = 8) fell in the *no school readiness screening* category. Table 2 displays additional descriptive information based on their screening procedures. Variance between the three groups was not statistically significantly different for any comparison. Using category (active, partial, or no school readiness screening) as the independent variable, no statistically significant differences were found between categories on number of districts [$F(2, 48) = 0.94$, $p = 0.40$], number of students enrolled [$F(2, 48) = 0.64$, $p = 0.53$], or annual median income [$F(2, 48) = 2.25$, $p = 0.12$].

Discussion

Transitioning into the school context is an essential developmental task of early childhood—one that requires navigating multiple cognitive, behavioral, and social-emotional challenges

simultaneously throughout the day (Sameroff & Haith, 1996). The context of formal schooling is qualitatively different from previous experiences children have had (Rimm-Kaufman & Pianta, 2000). Children accomplish this transition with more or less success, and the ease with which this transition occurs is indicative of success in future academic endeavors (e.g., Duncan et al., 2007; Graziano et al., 2007; Lonigan, 2006). Competence within certain developmental domains (e.g., cognitive, physical, social-emotional) is incorporated into the fundamental idea of "readiness" for school.

Despite much agreement on the importance of readiness for school entry, many children are reportedly "not ready" for this developmental challenge. In fact, a 2015 report from the U.S Department of Education indicated that 6 out of 10 preschool-aged children are not ready for kindergarten. Similarly, 46% of kindergarten teachers surveyed by the National Center for Early Development and Learning indicated that over half of the children in their class were not ready for school (Rimm-Kaufman, Pianta, & Cox, 2000).

Unfortunately, while on a national front there is movement toward consistency across the states in terms of academic criteria (e.g., Common Core), it is clear that this is not being done for the process of evaluating students' readiness to engage in these academic experiences. That is, no federal direction is currently offered to guide states and local agencies in their readiness strategies, nor to encourage these agencies to prioritize these efforts. It is not suggested that federal direction is simple, as indicated by challenges with Common Core, specifically related to common, standardized assessments (e.g., Jochim & McGuinn, 2016). It is difficult to find ways to have multiple agencies agree on priorities and how to spend resources. However, it is important to begin the dialogue on ways to enhance school readiness, something that has been determined to be vital to later success (Duncan, et al., 2007).

The results of this brief report indicate that just over half of the states and DC do in fact have active legislature documenting a state-wide school readiness screening practice. However, 16% of the states do not appear to have any formal school readiness screening practices in place. Based on student enrollment for these states, this leaves teachers, parents, and administrators of over 8 million students without a formalized readiness screening procedure. Furthermore, although 84% of states do appear to support some method of formalized school readiness screening (either full or partial), this information is often buried and difficult to find or the specifics of these practices are difficult to decipher, thus making it challenging to further support generalization. It should be noted that in the states where the authors were unable to find formalized procedures, this does not necessarily mean (a) that they do not exist, but rather that the current procedures for this search were unable to uncover them, and (b) that informal procedures do not exist.

Limitations

While concerted efforts were made to ensure each state was thoroughly investigated, there are clear limitations to this brief report. First, the search was limited to information that was readily available via a comprehensive search of State Administrative Rules, Administrative Codes, or similar databases, or information available on State DOE websites. It is possible that there are states actively making headway in the arena of school readiness screening, but that this information is not yet available to the public and thus not captured in this study. Second, it is possible that some of the data sources used for this study have not been updated, do not include complete school readiness screening information, or were updated following the data collection for this brief report, again providing a scenario in which the full picture of school readiness screening practices would not be captured in this study.

Future Directions

Despite these limitations, some comments about the ease with which screening practices can be found at the state level can be made. However, these policies and practices are not evaluated in terms of rigor or quality, but simply presence or absence. Future directions will clearly require these procedures to be evaluated on their merits, including reliability and validity of measures, timing of screening, reporting of these results to important stakeholders (e.g., parents, teachers), and interventions accessed if a student is deemed "not

Table 3

Sample Language for School Readiness Screening Policy

State	Policy Language	State Code
Delaware	"The Department shall adopt rules and regulations to implement a common statewide readiness tool that will review a child's readiness for learning when they enter kindergarten. The readiness tool shall serve as the basis for an objective readiness review conducted by the child's teacher or other members of the child's school team. The readiness tool shall review, but not be limited to, the following 5 domains: (1) Language and literacy development; (2) Cognition and general knowledge; (3) Approaches toward learning; (4) Physical well-being and motor development; and (5) Social and emotional development. (i) Implementation of the tool delineated in subsection (h) of this section above shall be phased in with the first identified kindergarten classes completing the readiness review in fall 2012. Thereafter the implementation of the readiness reviews shall be phased in with additional kindergarten classes participating in fall 2013 and fall 2014, with statewide implementation no later than fall 2015. The readiness reviews shall be completed within 30 school days of the start of school. A kindergarten student shall be required to be reviewed for readiness once during the student's enrollment in kindergarten. The Department regulations promulgated pursuant to this section shall address any exceptions to the requirement for implementation of the readiness tool for all students, based on factors such as a student's late enrollment in kindergarten."	Delaware Code Title 14 § 151 State Assessment System, Rules and Regulations (h-i)
Oregon	"(1) The Department of Education shall implement a kindergarten assessment as part of the statewide assessment system implemented pursuant to ORS 329.485. The kindergarten assessment shall allow for the assessment of children to determine their readiness for kindergarten. (2) The Department shall work jointly with the Early Learning Council to adopt a tool to be used for the kindergarten assessment. The kindergarten assessment shall measure areas of school readiness, which may include physical and social-emotional development, early literacy, language, cognitive (including mathematics), and logic and reasoning. The tool selected will be appropriate for all children including children with high needs and English language learners, and will align with Oregon's early learning and development standards as well as the adopted Common Core State Standards. (3) Prior to November 1, 2013 the department shall make the kindergarten assessment available to school districts. (4) Beginning with the 2013-2014 school year, all school districts shall administer the kindergarten assessment to students who are enrolled in kindergarten. (5) The Department shall include the results of the kindergarten assessment in the statewide longitudinal data system and shall provide the results of the kindergarten assessment to the Oregon Education Investment Board for inclusion in school district's achievement compacts."	Oregon Administrative Rule 581-022-2130 Kindergarten Assessment
Kentucky	"Section 2. Required Common Kindergarten Entry Screener. In accordance with KRS Chapter 45A, the Department shall adopt a statewide common kindergarten entry screener that: (1) Aligns with the definition of school readiness and the standards established in Building a Strong Foundation for School Success: Kentucky's Early Childhood Standards; (2) Assesses the domains of adaptive, cognitive, communication, motor, and social emotional as established in Building a Strong Foundation for School Success: Kentucky's Early Childhood Standards; (2) Assesses the domains cognitive, communication, motor, and emotional	704 Kentucky Administrative Rules 5:070. Common Kindergarten Entry Screener

> as established in Building a Strong Foundation for School Success: Kentucky's Early Childhood Standards; (3) Is a reliable and valid screener for its intended purposes for the target populations, including English learners and students with disabilities; and (4) Produces point-in-time student level results that indicate level of school readiness in the five (5) domains listed in subsection (2) of this section. Section 3. Administration of the Common Kindergarten Entry Screener. (1) Beginning in the 2013-2014 academic year, each Kentucky public school district shall administer the common kindergarten entry screener adopted by the Department in accordance with Section 2 of this administrative regulation. (2) Each school district shall administer the common kindergarten entry screener to each student entering kindergarten in the school district no earlier than fifteen (15) days prior to the start of the current academic year and no later than the thirtieth (30th) instructional day of the academic year. Section 4. Data Collection and Reporting. (1) Each school containing kindergarten students shall enter the data from the common kindergarten entry screener in the student information system within fifteen (15) instructional days of the district's administration of the common kindergarten entry screener but no later than October 15 of each academic year. (2) Data shall be reported by the Department at an aggregate level by: (a) School district; (b) School readiness domain; (c) Student demographics; and (d) Prior early learning settings. Section 5. District use of the common kindergarten entry screener. (1) A district shall provide individual student results of the screener to parents or guardians of individual students. (2) A district shall not use the common kindergarten entry screener results to determine eligibility for enrollment. All students who meet the enrollment requirements of KRS 159.010 shall be entitled to enter kindergarten without regard to the results of the common kindergarten entry screener. (3) Districts may use the common kindergarten entry screener data as a system of measurement in the following ways: (a) To inform districts, parents, and communities about early learning in order to close the school readiness gap; (b) To make informed policy decisions at the local level to support early learning experiences prior to school entry; (c) To establish local goals for program improvement in order to achieve early learning outcomes; and (d) To include data as evidences in the kindergarten through 3rd grade Program Evaluation under 703 KAR 5:230. (4) The results of the screener shall not be utilized as part of the school's or districts' overall score to determine recognition or support contained in any administrative regulation promulgated by the Board pursuant to KRS 158.6455."

ready." Finally, the impact these policies have on students will need to be evaluated to determine the extent that students benefit.

To conclude, having clear, consistent school readiness screening procedures seems imperative knowing how important this aspect of development is for youth. Additionally, while readiness has been linked to many important outcomes, school entry continues to be based solely on chronological age. With the tremendous amount of variation during this developmental period, this arbitrary criterion seems particularly problematic. However, a system solely based on school readiness screening is not without problems, either. For example, delaying school entry based on readiness screening may inadvertently penalize families who cannot afford to send their children to high-quality preschool for an additional year (Dockett & Perry, 2002). A "maturationist view" (i.e., the idea that children vary individually on readiness based solely on their developmental clock; Meisels, 1999) of delaying children who are deemed "not ready" would suggest that children will simply "become ready" when it is time, not influenced by high-quality experiences that they are having. Delaying entry to school is not the appropriate response to a

child not being ready based on screenings. Rather, consider the "social constructivist view" that children's readiness is embedded in their social and cultural context (Meisels, 1999). One should also recognize that screening does not simply replace the current method, but that it supplements it. When deficits at the time of screening are systematically identified, they can be planned for. High-quality and targeted activities and interventions can be implemented to augment the naturally occurring experiences in the Kindergarten classroom. It is hoped that with national direction, policy can follow to ensure that students are being served in the most optimal way, from school entry through graduation.

Considerations for Policy Reform

With school psychologists' training and expertise in data-based decision-making, experience in progress-monitoring efforts such as RtI and MTSS, and knowledge of the importance of evidence-based practices, school psychologists' participation in school readiness screening efforts seems to be a natural fit. There are several ways that school psychologists can get involved. Direct service at the school and district levels and with parents of the schools served is one way; advocacy is another. As emphasized by NASP (n.d.), school psychologists' participation in advocacy at the local, state, and federal level is essential not only to the profession, but also to ensure that all children have access to the services that aim to improve behavioral, social-emotional, and mental health outcomes. School psychologists can initiate conversations about school readiness screening and get involved in efforts to develop policy at the district and state levels.

At a local level, if a district does not already have a school readiness screening practice in place, school psychologists can advocate for the implementation of screening during the transition into kindergarten or early in the kindergarten year. Over time this screening practice can be turned into formal district policy. School readiness screening can even be linked to an RtI or MTSS model, if such a model exists in the district, to further enhance the utility of the tool (Diamond et al., 2016). Many states have developed their own school readiness screeners (e.g., Delaware Early Learner Survey, Florida Kindergarten Readiness Screener, Ohio's Kindergarten Readiness Assessment), while other states use tools already available including Teaching Strategies Gold (https://teachingstrategies.com) or DIBELS (https://dibels.uoregon.edu). Other screeners such as the Kindergarten Student Entrance Profile are free to use and have strong psychometric properties (Lilles, et al., 2009; see http://www.michaelfurlong.info/KSEP_2014/ for screener and resources).

On a larger scale, school psychologists could consider teaming up with their state school psychology organization (or similar organization), to advocate for policy change at the state level. When developing policy for school readiness screening consider including who will be screened (e.g., all kindergarten students, students new to the district in first grade), a timeframe for the screening (e.g., screening completed by November first, within thirty days of enrollment with the district), and domains to be screened (e.g., cognitive, academic, social-emotional, language). See Table 3 for sample language for such policy.

While it may seem daunting to "add another thing to the plate" of duties, it is possible that by attending to children's needs in the transition from preschool to kindergarten, it will ultimately reduce the services children will need later on.

References

Bowman, B., Donovan, M., & Burns, S. (Eds.), & the Committee on Early Childhood Pedagogy of the National Research Council. (2000). *Eager to learn: Educating our preschoolers.* Washington, DC: National Academy Press.

Bridgeland, J., DiJulio, J., & Morison, K. B. (2006). *The silent epidemic: Perspectives of high school dropouts.* Washington, DC: Civic Enterprises, LLC.

Canter, A. (2006). Problem solving and RTI: New roles for school psychologists. *Communiqué, 34* (5), insert. Retrieved from nasponline.org

Center for Public Education. (n. d.). How much money does our school district receive from federal, state, and local sources? Retrieved from www.data-first.org/data/how-much-money-does-our-schooldistrict-receive-from-federal-state-and-local-sources/

Condron, D. J. & Roscigno, V. J. (2003). Disparities within: Unequal spending and achievement in urban school district. *Sociology of Education, 76*, 18-36.

Costenbader, V., Rohrer, A. M., Difonzo, N. (2000). Kindergarten screening: A survey of current practice. *Psychology in the Schools*, *37*, 323-332.

Daily, S., Halle, T., Burkhauser, M., & Child Trends. (2010). A review of school readiness practices in the States: Early learning guidelines and assessments. *Early Childhood Highlights,* 1.

Davies, S., Janus, M., Duku, E., & Gaskin, A. (2016). Using the Early Development Instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly, 35,* 63-75. doi: 10.1016j.ecresq.2015.10.002

Diamond, E., Quirk, M., & Furlong, M. (2016). Academically resilient Latino elementary students bridging the achievement gap. *Contemporary School Psychology, 20,* 160-169.

Diffey, L., Parker, E., & Atchison, B. (2017). State pre-K funding 2016-17 fiscal year: Trends and opportunities. Retrieved from https://www.ecs.org/ec-content/uploads/State-Pre-K-Funding-2016-17-Fiscal-Year-Trends-and-opportunities-1.pdf

Dockett, S., & Perry, B. (2002). Who's ready for what? Young children starting school. *Contemporary Issues in Early Childhood, 3*, 67-89. doi: 10.2304/ciec.2002.3.1.9

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428-1446

Florida Office of Early Learning. (2017). http://www.floridaearlylearning.com/home.aspx

Graziano, P. A., Reavis, R. D., Keane, S. P., & Calkins, S. D. (2007). The role of emotion regulation in children's early academic success. *Journal of School Psychology*, *45,* 3–19. doi:10.1016/j.jsp.2006.09.002

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). An exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher,23,* 5-14.

Hojnoski, R. L., & Missall, K. N. (2006). Addressing school readiness: Expanding school psychology in early education. *School Psychology Review, 35*, 602-614.

Jochim, A., & McGuinn, P. (2016, Fall). The politics of the Common Core assessments: Why states are quitting the PARCC and Smarter Balanced testing consortia. *Education Next, 16*. Retrieved from http://educationnext.org/the-politics-of-common-core-assessments-parcc-smarter-balanced/

Karoly, L. A., Kilburn, M. R., & Cannon, J. (2005). *Early childhood interventions proven results, future promise.* Santa Monica, CA: The RAND Corporation.

Lilles, E., Furlong, M., Quirk, M., Felix, E., Dominguez, K., & Anderson, M. (2009). Preliminary development of the Kindergarten Student Entrance Profile. *The California School Psychologist, 14,* 71–80. doi:10.1007/BF03340952

Lonigan, C. (2006). Development, assessment, and promotion of preliteracy skills. *Early Education and Development, 17,* 91–114. doi:10.1207/s15566935eed1701_5

Meisels, S. J. (1999). Assessing readiness. In R. C. Pianta & M. J. Cox (Eds.), *The transition to kindergarten* (pp. 39-66). Baltimore, MD: Paul Brooks.

National Association of School Psychologists. (n.d.). Advocacy tools & resources. Retrieved from: nasponline.org/research-and-policy/advocacy-tools-and-resources

National Association of School Psychologists. (2010). NASP model for comprehensive and integrated school psychological services: NASP Practice Model overview. *Communiqué, 39*(4), insert. Retrieved from: nasponline.org.

National Education Association (2016). *Rankings and estimates: Rankings of the states 2015 and estimates of school statistics 2016*. Washington, DC: Author.

Rimm-Kaufman, K., & Pianta, R. C. (2000). An ecological perspective on the transition to kindergarten: A theoretical framework to guide empirical research. *Journal of Applied Developmental Psychology, 21*, 491-511. doi:10.1016/s0193-3973(00)00051-4

Rimm-Kaufman, K., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*, 147-166. doi: 10.1016/s0885-2006(00)00049-1

Rumberger, R. W., & Arellano, B. D. (2009). Understanding and addressing the California Latino achievement gap in early elementary school. In R. A. Gutiérrez & P. Zavella (Eds.), *Mexican Americans in California: Transformations and challenges* (pp. 61–76). Champaign, IL: University of Illinois.

Sameroff, A. J., & Haith, M. M. (1996). *The five to seven year shift: The age of reason and responsibility*. Chicago, IL: University of Chicago Press.

SPSS Inc. (2013). *IBM SPSS Statistics, version 22.0*. Chicago, IL: Author.

Thurlow, M. L., & Gilman, C. J. (1999). Issues and practices in the screening of preschool children. In E. V Nuttall, I. Romero, & J. Kalesnik (Eds.), *Assessing and screening preschoolers: Psychological and educational dimensions* (pp. 72-93). Austin, TX: Proed.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161–188). Mahwah, NJ: Erlbaum.

U.S. Department of Education. (2009). The Early Learning Challenge Fund: Results-oriented, standards reform of state early learning programs. Retrieved from http://www2.ed.gov/about/inits/ed/earlylearning/elcf-factsheet.html

U.S. Department of Education. (2015). A matter of equity: Preschool in America. Retrieved from https://www2.ed.gov/documents/early-learning/matter-equity-preschool-america.pdf

U.S. Department of Education. (2016). Preschool development grants. Retrieved from https://www2.ed.gov/programs/preschooldevelopmentgrants/index.html?src=rotator

Research and Practice in the Schools
2018, Vol. 5, No. 1, pp. 29-37

ISSN: 2329-5783

Article

Kindergarten-Monitoring Instructional Responsiveness: Reading (K-MIR:R): Examination of an Authentic Curriculum-Based Measure of Beginning Reading Skills

Kathleen B. Aspiranti
Youngstown State University

Angela Hilton-Prillhart
Milligan College

Sherry M. Bell and R. Steve McCallum
University of Tennessee, Knoxville

The Kindergarten-Monitoring Instructional Responsiveness: Reading (K-MIR:R; Bell, Hilton-Prillhart, McCallum, & Hopkins, 2011), a brief (3-minute) experimental group-administered early reading screener is designed to function as a curriculum-based measure that assesses three critical beginning reading skills. The K-MIR:R was administered to 99 students yielding strong test-retest reliability (r = .84) and statistically significant inter-correlations across scales (p < .01: Letter Matching, Sight-Word Identification, and Comprehension). Concurrent validity was investigated by comparing scores on two administrations of K-MIR:R probes to the Discovery Education Assessment (DEA; Discovery Education, 2013) scores; coefficients were .59 and .64 respectively. In order to examine how well each of the probes predicted DEA classifications of proficient and below proficient, a Discriminant Function Analysis was conducted. Form A_1 accurately predicted 90.4% of the cases and Form A_2 accurately predicted 80.9%. Implications for further research are discussed.

Key words: Reading, assessment, curriculum based measurement, kindergarten

The importance of accurate reading screening leading to early intervention with kindergarten students is well-documented (Invernizzi, Justice, Landrum, & Booker, 2004; Rathvon, 2004; Snow, Burns, & Griffin, 1998). However, many of the currently available screeners have limitations (e.g., they require individual administration, are lengthy, and/or exhibit insufficient validity evidence). The purpose of this study is to examine the reliability and concurrent and predictive validity of an experimental multidimensional kindergarten reading screener delivered in a brief, time-limited, group-administered format as a potential first step in a universal screening process.

According to the literature, multifaceted screeners that include phonemic awareness, letter knowledge, and expressive and receptive vocabulary may be the most effective way to correctly determine the reading performance of young students (Compton, Fuchs, Fuchs, & Bryant, 2006; Jenkins & O'Connor, 2002; McCardle, Scarborough, & Catts, 2001; O'Connor & Jenkins, 1999; Scarborough, 1998; Torgesen, 2002). Typically, assessment of phonemic awareness is conducted in first grade along with decoding, word identification, and text reading (Foorman et al., 1998). O'Connor and Jenkins (1999) found that the most effective kindergarten screeners were multifaceted measures containing letter naming fluency, letter sound identification, blending onset and rimes, phoneme segmentation, and sound repetition. Additionally, pseudoword reading in

young children may be an important predictor of reading difficulties as well as a predictor of students who will gain the most from phonics-based interventions (Van der Kleij, Segers, Groen, & Verhoeven, 2017).

Many educators have expressed concern about instructional time lost due to conducting assessment testing (Nelson, 2013; Valli & Buese, 2007). One solution is to use group-administered assessments, which may be appealing to teachers because of the efficiency and ease of administration. Some group-administered screening measures designed for use in kindergarten are available. Tests commonly used for kindergarten reading screening include the Basic Early Assessment of Reading (BEAR; Riverside, 2002), Group Reading Assessment and Diagnostic Evaluation (GRADE; Pearson, 2010), and the Predictive Assessment of Reading (PAR; Wood, 2001-2013). The BEAR and GRADE are somewhat time intensive to administer; the BEAR requires up to 45 minutes to administer and the GRADE can take as long as 50-90 minutes. Skills measured on the BEAR include phonemic awareness, phonics, vocabulary, comprehension, and oral reading fluency, which requires individual administration. The GRADE measures performance on phonological awareness, vocabulary, reading comprehension, and listening comprehension. The PAR is more time-efficient, taking as little as 16 minutes to administer; however, the PAR is individually administered in a computerized format. The PAR assesses letter-word calling, picture name vocabulary, phonological awareness, and rapid naming fluency.

Like the PAR, there are several other assessments that are administered to students on individual computers. Although each student takes the test separately and at their own pace, the test can be administered to many students at once. Examples of these types of reading assessments are i-Ready, Measures of Academic Progress (MAP), and STAR. I-Ready is a computerized reading screener that takes between 30 and 60 minutes to complete and is computer scored. This adaptive test uses the Common Core State Standards (CCSS) to measure phonological awareness, phonics, high-frequency words, vocabulary, comprehension of information text, and comprehension of literature (Curriculum Associates, nd). The MAP is a 40-minute, computer-adaptive test based on CCSS. The primary version (grades K-1) measures foundational skills, language and writing, literature and informational text, and vocabulary use (Northwest Evaluation Association, 2017). STAR is another computer-adaptive test based on CCSS; the early literacy version (grades PK-3) takes approximately 20 minutes to complete. The STAR assesses word facility and skills, comprehension strategies and constructing meaning, and numbers and operations (Renaissance Learning, 2015). Another instrument commonly used for screening purposes in early grades is the Discovery Education Assessment (DEA), a computer administered group assessment aligned to common core standards (Discovery Education, 2013). The DEA takes approximately 40 minutes to administer and assesses literature, information, foundations, writing, language, and speaking/listening. For kindergarten students, the assessment consists of 28 items.

The purpose of the current study is to explore the psychometric qualities of a newly developed time-limited, group-administered screener, and more specifically determine its efficacy in assessing the early reading skills of kindergarten students as a first step in a Response to Intervention (RtI) screening model. The Kindergarten-Monitoring Instructional Responsiveness: Reading (K-MIR:R) is part of a group-administered universal screening and progress monitoring package of assessments of reading for grades K-3; multiple equivalent forms probes assess reading at grade levels K through 3 (Bell, Hilton-Prillhart, McCallum, & Hopkins, 2011; Hilton-Prillhart, 2011). The MIR:R (formerly called Monitoring Academic Progress-Reading or MAP-R) for grades 1-3 is a silent, group-administered screener which was piloted as part of a comprehensive RtI program. MIR:R along with AIMSweb© Maze and STAR reading were administered to 1,688 students in Grades 1-3 (Hilton-Prillhart, 2011). Overall alternate-form reliabilities for MIR:R resulted in moderately high stability (Grade 1 =.79, Grade 2 = .78, and Grade 3 = .75). Test-retest reliability was .90 for Grade 1, .84 for Grade 2, and .89 for Grade 3. With regard to concurrent validity, correlations for MIR:R and AIMSweb© Maze ranged from .43 to .69, with correlations for MIR:R and STAR ranging from .48 to .67. Predictive validity was assessed using end-

of-the-year STAR reading scores as the criterion for MIR:R and AIMSweb© Maze. Results of a stepwise regression indicated that MIR:R scores predicted 37% of the variance in STAR scores and AIMSweb© Maze failed to add additional predictive variance. Data support the utility of MIR:R as a reading screener for progress monitoring within a RtI framework (Hilton-Prillhart, 2011).

The K-MIR:R assesses skills related to letter knowledge and phonological awareness, as well as basic sight word proficiency and comprehension. The K-MIR:R's inclusion of an array of reading skills is in line with research that suggests as children prepare to move into grade 1, screeners should include measures of phonemic awareness, decoding, and word reading with speed as an outcome (Compton et al., 2006; Fuchs, Fuchs, & Compton, 2004; O'Connor & Jenkins, 1999). The K-MIR:R also endeavors to identify not only children who are in need of additional screening for intervention, but also those students with advanced skills. The current study adds to the body of literature by examining the reliability and validity of a new brief, group-administered assessment that assesses the letter knowledge, phonological awareness, sight word reading, and comprehension skills necessary for early reading success. Specific research questions for this pilot study include: 1) To what extent does the K-MIR:R assessment exhibit test-retest reliability?; 2) What is the concurrent validity of the K-MIR:R when compared to another well-accepted measure?; 3) Do student scores on the K-MIR:R correctly predict classification into either proficient or below proficient categories on the DEA?

Method

Participants

Data were collected from 99 kindergarten students (44 females and 55 males) from two elementary schools in a rural Southeastern school district. Both schools were designated as Title I schools with Free/Reduced Lunch rates ranging from 72% to 93%. The average age of the students was 5.6 years (*SD* = .56); 13 students had Individualized Educational Plans (IEPs) (11 for speech/language disability, two for developmental delays), and five were receiving English as a Second Language (ESL) services. Reported race of participants was: Caucasian (77%), Hispanic (12%), African American (5%), Asian/Pacific Islander (1%), and not identified (5%).

Instruments

K-MIR:R. K-MIR:R probes are designed to be used for universal screening and progress monitoring, with multiple alternative forms. Key features of the MIR instruments are that they measure more than one skill and they are group-administered; the goal is to yield valid information in a timely manner. A classroom of up to 30 students can be assessed at once with a 3-minute screener. Traditional Curriculum-Based Measurements (CBM) typically assess only one skill and/or require time-intensive individual administration. K-MIR:R probes contain 35 items per probe, yielding a possible range of scores from 0 to 35. Each item presents the student with a stimulus followed by three options. Using scripted instructions, examinees are instructed to select the best answer choice for each item. Words are written in Times-New Roman 14-point font and all are presented in lower-case with equal numbers of spaces between each letter. The probes are group-administered in a paper format and students are given 3 minutes to complete each probe, after completing a guided practice exercise that exposes them to each item type. The practice exercise takes approximately 4 minutes to complete prior to administering the K-MIR:R.

The K-MIR:R measures four basic reading abilities that increase in difficulty: Letter Matching (LM; eight items), Decoding (D; eight items), Sight Word Reading (SR; eight items), and Comprehension (C; seven items). Items also exist that measure D, SR, and C at the first-grade level to allow for differentiation of high-achieving students. Letter Matching consists of matching an upper-case letter to a lower-case letter. Decoding consists of identification of initial and final consonants and medial vowel sounds. For Decoding, the student is to select the word that best matches a picture prompt from three word choices (e.g., the item shows a picture of a hat and the student is to choose the best match from *hot, hat*, and *hit*). Sight-Word Reading requires students to match a picture to the

correct sight word from three response options. Comprehension requires students to examine a picture, then determine (or infer) the word that best matches the picture (e.g., a picture of a steaming cup of coffee is best matched to the word *hot*).

The items on the K-MIR:R were reviewed by two experts in early reading content to examine the content validity of the items. Items were added or deleted as necessary. To ensure adequate floor and ceiling, the skill level of the items is a mix of skills expected at kindergarten through first grade, as determined by use of sight words from the Dolch word lists and recommendations for skills based on the Common Core State Standards (National Governors Association Center for Best Practices, 2010). Item difficulty is rotated and each form contains an equal amount of item types. K-MIR:R probes are easily and quickly scored (15-30 seconds per probe) using a scoring template that is a replica of the probe with the correct item indicated by a slash mark.

DEA. The DEA is a standardized, computerized, standards-based universal assessment that all students in the participating school district take three times a year as a benchmark assessment. The DEA is administered in a group setting with the students at individual computers with headphones. The items are displayed one at a time on the computer screen, and the student uses a mouse to click on an answer choice. At the kindergarten level, the items are read aloud through the headphones. The DEA assesses standards in the areas of literature, information, foundations, writing, language, and speaking/listening. For kindergarten students, the assessment consists of 28 items. The skills of phonological awareness and phonics fall under the foundations category, which has a relatively small sample of eight items. Administration of the DEA is not timed but takes approximately 40 minutes for students to complete. The test is then automatically scored by the computer, providing scaled scores, percentiles, and descriptive bands according to the percent of items answered correctly.

The DEA is reportedly aligned to Common Core standards with reliability for the standardization sample ranging from r = .73 to .89 (Discovery Education, 2013). Additionally, the concurrent validity with the Kentucky KCCT State Mandated Test (KY) and the DEA ranged from r = .48 to .72; Concurrent validity was r =.58 to .72 between the DC CAS State Mandated Test (DC) and DEA. Predictive validity between the DEA and KY ranged from r = .61 to .72 and from r = .44 to .74 between DEA and DC (Center on Response to Intervention, 2014).

The National Center's Technical Review Committee on Screening used a rating scale to examine the evidence for screening tools. Classification accuracy, or the ability to accurately classify students into at-risk categories, used area under the curve (AUC) and four questions (see Center on Response to Intervention, 2014) to identify convincing, partially convincing, or unconvincing evidence. Partially convincing evidence required AUC > .85 and three of the four questions rated as Yes. Validity measures used three questions: Was convincing evidence supporting content validity presented?; Was convincing construct validity presented with correlations above .70?; Was convincing predictive validity presented with correlations above .70? Partially convincing evidence required two questions rated as Yes, while convincing evidence required all three questions answered Yes. The DEA had convincing evidence for reliability and classification accuracy, moderately high generalizability, and partially convincing evidence for validity (Center on Response to Intervention, 2014; Discovery Education Predictive Assessment, 2014). When these scores are compared to other reading screeners, AimsWeb reading CBM has partially convincing evidence for classification accuracy and validity; AimsWeb letter naming fluency has moderately low generalizability and partially convincing evidence for validity; GRADE has moderately low generalizability and partially convincing evidence for classification accuracy and validity; MAP has partially convincing evidence for reliability and validity; PALS early literacy has partially convincing evidence for validity; and STAR early literacy has partially convincing evidence for classification accuracy and validity (Center on Response to Intervention, 2014). Therefore, the DEA is a comparable measure to other well-accepted assessment measures (Center on Response to Intervention, 2014).

Procedures

Permission was obtained from school principals, teachers and parents in accordance with University guidelines on rights of human subjects before beginning the K-MIR:R administration. The K-MIR:R (FormA_1) was administered class-wide to 78 students in kindergarten classrooms by one of the district's school psychologists and an advanced graduate student in mid-April. All directions were read aloud by the school psychologist. The three-page K-MIR:R was distributed with a practice sheet on top. The practice sheet contained four problems that represent the different types of items presented in the K-MIR:R. Students completed the problems on the practice sheet at the direction of the school psychologist. Each practice item was explained and checked for completeness and accuracy before moving to the next practice item. Students were then asked if they had any questions. After ensuring that the task was understood, students were instructed to look at the three page K-MIR:R booklet and to begin. Two weeks later, to examine test-retest reliability, the same probe (labeled Form A_2) was administered following the same procedures as above to a total of 91 students; several were out of their classrooms for the administration of Form A_1. One week later (in late April), the DEA was administered via computer to 99 students by classroom teachers as part of district-wide assessment practices. All students administered the DEA took at least one form of the K-MIR:R assessment. The DEA was scored via computer per the process prescribed by DEA and adopted by the participating school district. K-MIR:R Forms A_1 and A_2 were scored by the school psychologist using a scoring key. Interrater reliability was obtained by comparing scores from a trained graduate student who scored 22% of the probes (n=37). Scores for each item were compared across the two raters on an item-by-item basis. Cohen's Kappa indicated a high rate of agreement between raters, Kappa = 0.96 ($p < .001$).

Results

Means, ranks, skewness, and kurtosis for each administration of the K-MIR:R, designated Form A_1 and A_2, respectively, are shown in Table 1. Participants' mean scores on the K-MIR:R were 12.33 for Form A_1 and 15.00 for Form A_2. It was clear to the administrators that some of the students did not understand the directions and were unable to complete the probes. Consequently, three students who earned total scores of less than 3 were considered "outliers," and their scores were excluded from further analyses (for a rationale see Miller, Bell, & McCallum, 2015; Osborne & Overbay, 2008). See Table 1 for the range of scores, skewness, and kurtosis, which are acceptable for K-MIR and DEA scores (i.e., < 1).

The test-retest correlation between Form A_1 and Form A_2 was .84, $p < .001$ ($n = 78$). Correlation coefficients among individual item types were obtained across probes as shown in Table 2. The correlation between Decoding, Sight Word Recognition, Comprehension, and First Grade Skills item types were all statistically significant, ranging from .25 to .50. Letter Matching scores did not significantly correlate between Form A_1 and A_2.

To determine concurrent validity, scores on the K-MIR:R probes were compared to student spring DEA benchmark scores; moderately strong correlation coefficients between Form A_1 and A_2 and the DEA were obtained ($r = .59$, $p < .001$, $n = 78$; $r = .64$, $p < .001$, $n = 91$, respectively). These correlations provide some evidence of concurrent validity for the K-MIR:R when compared to a widely accepted kindergarten reading screener. Individual item types on each probe were also compared to DEA scores. On Form A_1, Letter Matching, Sight Word Recognition, Comprehension, and First Grade Skills all significantly correlated with DEA scores, with correlations ranging from .28 to .34. On Form A_2, Letter Matching, Comprehension, and First Grade Skills all significantly correlated with DEA scores, with correlations ranging from .36 to .44.

In order to determine the extent to which K-MIR:R probes predict student performance on the DEA, a Discriminant Function Analysis (DFA) was conducted. DFA is considered a reasonable analysis when one or more continuous variables are used to predict a categorical variable, and is considered more appropriate than logistic regression when the sample size is small and outliers are eliminated (Beleites et al., 2011). For the DFA analyses, one predictor score was generated by combining all

Table 1

Means, Standard Deviations, and Score Bands for Kindergarten-Monitoring Instructional Responsiveness-Reading (K-MIR:R) Probes and Discovery Education Assessment (DEA)

Probe	*N*	*M (SD)*	Minimum Score	Maximum Score	Skewness (Std. Error)	Kurtosis (Std. Error)
K-MIR:R Form A_1	78	12.33 (6.45)	3	32	.78 (.27)	-.29 (.54)
K-MIR:R Form A_2	91	15.00 (7.70)	3	34	.47 (.25)	-.61 (.50)
DEA	99	1256 (59.08)	1140	1375	.65 (.25)	-.73 (.49)

Table 2

Test-Retest Correlations between Kindergarten-Monitoring Instructional Responsiveness: Reading (K-MIR:R) Probe Item Type Subscales for Form A_1 and Form A_2 and Discovery Education Assessment (DEA) Score

	A_1 (LM)	A_2 (LM)	A_1 (D)	A_2 (D)	A_1 (SR)	A_2 (SR)	A_1 (C)	A_2 (C)	A_1 (F)	A_2 (F)
A_1 (LM)	--									
A_2 (LM)	-.06	--								
A_1 (D)	-.03	.66**	--							
A_2 (D)	-.02	-.02	.25*	--						
A_1 (SR)	.07	.22*	.18	.10	--					
A_2 (SR)	.05	.11	-.09	.26*	.50**	--				
A_1 (C)	.34**	-.15	.21	.10	.30*	.36**	--			
A_2 (C)	.14	.28**	.21	.20	.36**	.15	.50**	--		
A_1 (F)	.33**	-.04	.10	-.04	.30**	.36**	.42**	.28*	--	
A_2 (F)	.20	.15	.18	.05	.41**	.31**	.31**	.28**	.45**	--
DEA Score	.32**	.40**	.10	.12	.27*	.20	.28*	.44**	.34**	.36**

Note: * $p < .05$; ** $p < .001$; LM = Letter Matching, D = Decoding, SR = Sight Word Reading, C = Comprehension, F = First Grade Skills

Table 3

Classification Results Table for Predicted Group Membership in Discovery Education Assessment (DEA) Proficient/Below Proficient Categories based on Combined Item Type Scores on Kindergarten-Monitoring Instructional Responsiveness: Reading (K-MIR:R) Form A1 and Form A2

			Predicted Group Membership		
			Below Proficient	Proficient	Total
Form A_1*	Original Count	Below Proficient	7	1	8
		Proficient	6	59	65
Form A_2**	Original Count	Below Proficient	10	6	16
		Proficient	11	62	73

*90.4% of original grouped cases correctly classified
**80.9% of original grouped cases correctly classified

probe scores (to form an early reading skills operationalization). The DEA provides four categories of scores (below basic, basic, proficient, and advanced). For this study, the four categories were collapsed into two (below proficient and proficient) because the goal is for students to achieve at least a proficient level in reading. For Form A_1, the Chi-square was significant (Wilks Lambda = .67, Chi-square = 27.27, *df* = 5, Canonical correlation = .57, $p < .001$), and the classification results in Table 3 reveal that 90.4% of students were accurately classified into the two DEA categories by the combined item types from Form A_1. Further, the DFA displayed evidence that accuracy for prediction into the proficient category (90.8%) was better than for the prediction into the below proficient category (87.5%).

For Form A_2, the Chi-square was also statistically significant (Wilks Lambda = .65, Chi-square = 36.14, *df* = 5, Canonical correlation = .59, $p < .001$). The classification results in Table 3reveal that 80.9% of the students were classified into the two DEA proficiency categories by the combined item types from Form A_2. Proficient scores were more accurately classified (84.9%) than below proficient scores (62.5%).

Discussion

The purpose of this study was to establish preliminary support for the K-MIR:R as a group administered screener of Kindergarten reading skills. Results modestly indicate that the K-MIR:R has promising basic psychometric properties (i.e., acceptable skewness and kurtosis estimates and strong test-retest reliability). Additional evidence of test-retest reliability is provided by the significant correlation coefficients between the subscales of the probes, with one exception. Form A_1 and Form A_2 Letter Matching scores were not significantly correlated, potentially due to the unexpected difficulty of one item within this subscale that required matching an upper case 'B' to lower case 'b' with 'p' and 'd' as distracters. More than 25% of the students answered this item incorrectly whereas on the other Letter Matching items the percent of students answering each item incorrectly was less than 10%.

Results also provide evidence of relatively strong concurrent validity of the K-MIR:R when compared with the DEA. The 3-minute group K-MIR:R correlates moderately with the DEA, a 40-minute, computer-administered test. These results suggest that K-MIR:R has the potential to provide valuable screening data, an important characterization given its efficient structure (i.e., brief, group administered, pencil-paper format). The case for using K-MIR:R is enhanced as a function of two developmental characteristics. Specifically, it was designed as a curriculum-based measure. Even though administration time is brief (3 minutes), the K-MIR:R is unique in that it measures more than one skill. This multidimensionality is consistent with reading experts' assertions that several early reading skills are important building blocks of literacy and should be assessed in kindergarten and first grade (e.g., Compton et al., 2006; McCardle et al., 2001; Torgesen, 2002).

The mostly moderate correlation coefficients among items on the separate probe types (e.g., Letter Matching, Decoding, Sight Word Reading, Comprehension, First-Grade Skills) and the significant correlations between three of the five probe types and the DEA scores suggest that there is some overlap in the skills measured by the different probe types as shown in Table 2. More importantly, these results provide evidence for semi-independence of the subscales. That is, scores from each probe type reflect a somewhat independent but ultimately important literacy skill. The importance of assessing several variables of early reading is needed for a kindergarten screener, as the most effective kindergarten screeners are multifaceted (O'Connor & Jenkins, 1999).

Evidence for the predictive value of the K-MIR:R is reflected by results from the Discriminant Function Analyses. The K-MIR:R scores provide accurate classification of students as either proficient or below proficient on the DEA. True positives, true negatives and related classifications (false positive/negatives) demonstrate prediction, as shown in Table 3. The first administration, Form A_1, provided slightly better predictive power. That is, for Form A_1 only one student (12.5%) was identified as a false negative. However, on Form A_2 six students (37.5%) were identified as false negatives, but the difference was not significant. Perhaps the practice effect influenced performance

on Form A_2 negatively. Those students who were truly lower performers may perform better on the K-MIR:R with practice and repeated assessments.

Limitations and Future Research

This study had several limitations. For example, participants came from elementary schools in one rural school district in one geographic location, which may not adequately reflect the general population of all kindergarten students. Additionally, students were allowed to skip items to ensure they were able to answer all items with which they were familiar. This procedure could have led to some item types being less represented than others in the analyses. Finally, data were collected near the end of the school year rather than the beginning, which is more typical for a screener. Additional research is needed to examine the psychometric properties of the K-MIR:R (e.g., by comparing K-MIR:R scores to other instruments, evaluating the utility in other locales and types of school districts). Though the K-MIR:R is designed as a multidimensional measure, the current study was not designed to examine the viability of subscales as independent and valid constructs. A larger sample size and different administration instructions (i.e., prohibiting skipping items) will be needed to examine the validity of the subscales.

In summary, this study adds to the literature by demonstrating the promise of a group-administered, multidimensional reading screener for Kindergarten students. An accurate screener that takes little time from classroom instruction is potentially quite worthwhile. The K-MIR:R fulfills several criteria for an effective reading screener for kindergarten students. First, it incorporates multifaceted item types containing letter matching, letter sound identification (embedded in decoding), and onset and rimes (embedded in decoding) critical for early reading success (O'Connor & Jenkins, 1999). Second, it is group-administered, significantly decreasing the amount of time a teacher needs to take from instruction for assessment. Finally, the K-MIR:R is time efficient, user-friendly, and inexpensive, taking only 5 to 6 minutes total to complete both the practice and administration using a pencil-paper format.

References

Beleites, C., Geiger, K., Kirsch, M., Sobotika, S. B.,Schackert, G. S., & Salzer, R. (2011). Raman spectroscopic grading of astrocytoma tissues: Using soft reference information. *Analytical and Bioanalytical Chemistry, 400,* 2801-2816.

Bell, S. M., Hilton-Prillhart, A., McCallum, R. S., & Hopkins, M. (2011). *Monitoring Instructional Responsiveness: Reading (MIR:R).* (Unpublished test). University of Tennessee, Knoxville.

Center on Response to Intervention. (2014). *Screening Tools Chart.* Washington, DC: American Institute for Research.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*, 394-409.

Curriculum Associates. (nd). *The Science Behind i-Ready's Adaptive Diagnostic.* North Billerica, MA: Curriculum Associates.

Discovery Education. (2013). *Discovery Education Assessment*. Nashville, TN: Discovery Communications, LLC.

Discovery Education Prediction Assessment. (2014). Washington, DC: US Department of Education, Office of Special Education Programs, Center on Response to Intervention at American Institutes for Research. Retrieved From http://www.rti4success.org/discovery-education-predictive-assessment-reading.

Foorman, B. R., Fletcher, J. M, Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A.…Taylor, R. H. (1998). *Technical Report Texas Primary Reading Inventory* (1998 Edition). Houston, TX: Center for Academic and Reading Skills and University of Houston.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children*, *71*(1), 7–21.

Hilton-Prillhart, A. (2011). *Validation of the Monitoring Instructional Responsiveness: Reading (K-MIR:R): Development and investigation of a group-administered comprehension-based tool for RTI.* (Unpublished doctoral dissertation). University of Tennessee, Knoxville.

Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36,* 479-500.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. C. Danielson, & D. P. Hallahan (Eds.), *Identification of Learning Disabilities: Research to Practice* (pp. 99-149). Mahwah,NJ: Erlbaum.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disability Research and Practice, 16*, 230-239.

National Governors Association Center for Best Practices. (2010). *Common core state standards.* National Governors Association Center for Best Practices, Washington, D.C.

Nelson, H. (2013). *Testing more, teaching less: What America's obsession with student testing costs in money and lost instructional time.* New York: American Federation of Teachers.

Northwest Evaluation Association. (2017). *MAP Skills.* Portland, OR: Northwest Evaluation Association.

O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159–197.

Osborne, J. W., & Overbay, A. (2008). Best practices in data cleaning. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 205-212). Thousand Oaks, CA: Sage Publications.

Pearson. (2010). *Group Reading Assessment and Diagnostic Evaluation*. New York, NY: Pearson.

Rathvon, N. (2004). *Early reading assessment: A Practitioner's handbook.* New York: Guilford Press.

Renaissance Learning (2015). *Parent's Guide to STAR Assessments.* Wisconsin Rapids, WI: Renaissance Learning.

Riverside Publishing Company (2002). *Basic Early Assessment of Reading*. Itasca, IL: Riverside Publishing Company.

Scarborough, H. S. (1998). Early identification of children at risk for reading difficulties: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–199). Timonium, MD: York Press.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). P*reventing reading difficulties in young children. Committee on the Prevention of Reading Difficulties in Young Children, National Research Council*. Washington, DC: National Academy Press.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7-26.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal, 44,* 519-558.

Van der Kleij, S. W., Segers, E., Groen, M. A., & Verhoeven, L. (2017). Response to intervention as a predictor of long-term outcomes in children with dyslexia. *Dyslexia, 23,* 268-282.

Wood, F. (2001-2013). *Predictive Assessment of Reading.* Clemmons, NC: Red-e Set Grow.