Received: 23 January 2020 | Revised: 29 December 2020 | Accepted: 5 January 2021

DOI: 10.1002/pits.22480

RESEARCH ARTICLE

WILEY

Current status and future directions in assessment of paraprofessional practices

Adam J. Lekwa | Linda A. Reddy

Department of School Psychology, Graduate School of Applied and Professional Psychology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

Correspondence
Adam J. Lekwa, Department of School Psychology, Graduate School of Applied and Professional Psychology, Rutgers, The State University of New Jersey, 41 Gordon Rd, Suite C, Piscataway, NJ 08854, USA.
Email: adam.lekwa@rutgers.edu

Funding information
Institute of Education Sciences, Grant/Award Number: R324A170069

Abstract

Paraprofessionals have become a significant portion of the educational workforce for students with or at risk for needing, or receiving special educational services. Although best practice and federal legislation require supervision of paraprofessionals by certified professional educators, research indicates that supervision rendered to paraprofessionals is often ad hoc, infrequent, and not based on data. We critically examined the availability and validity evidence of existing assessment approaches (published or publically posted) for evaluation and monitoring of paraprofessional practices in schools in the United States. Overall results from a systematic review of research literature and a review of materials made available publically by schools yielded 14 assessment tools with a focus on 18 constructs. Each assessment lacked reliability and validity evidence, and had limited utility for guiding paraprofessional supervision and professional development. Based on the larger paraprofessional literature and availability of assessments, a framework of recommendations for measurement development and validation is proposed. Specifically, we offer suggestions for domains of practice and validation efforts for evaluating forms of reliability, construct validity, and consequential validity for new assessment approaches for these important school personnel.

KEYWORDS
evaluation, formative assessment, paraprofessional, professional development

Students with the greatest need are increasingly served by school staff (i.e., paraprofessionals) with the least training and in-service professional development or guidance. There are approximately 1 million paraprofessionals serving students in US schools (e.g., Dulfer, 2013; U.S. Department of Labor Bureau of Labor Statistics, 2016); most work with students with disabilities, or students at risk for requiring special education (e.g., Blalock, 1991; Carter et al., 2009; Scull & Winkler, 2011). While the minimum qualifications for paraprofessionals vary by state, many states require a high school diploma while other states require associates' degrees or equivalent training. Estimates indicate that approximately 40% of states have hired more full-time paraprofessionals than full-time special educators (U.S. Department of Education, 2017). In light of these trends nationwide, concern surrounding the effectiveness of the services offered by paraprofessionals, and the adequacy of the supervision and training they receive have been increasing (Farrell et al., 2010). Despite this increasing level of concern, research on the assessment or evaluation of paraprofessional practices or effectiveness appears to be minimal. The purpose of this study was to take stock of assessment approaches designed to measure or evaluate the practices of paraprofessionals and current school-based practices to inform progress in the development of assessments in this domain.

1 | PARAPROFESSIONAL ROLES, SUPERVISION, AND DEVELOPMENT

Paraprofessionals are part-time or full-time school personnel who assist classroom teachers in supporting the learning and behavior of students with or at-risk special needs. Any assessment of paraprofessional practices must be designed to account for specific paraprofessional roles, needs related to successful supervision, and generation of feedback to promote the development of skills in support of student functioning and learning.

Paraprofessionals, historically, were expected to perform nonspecialized clerical tasks such that teachers' time for instruction might be less restricted. Employment of paraprofessionals over time underwent a massive expansion, and so, too, have the functions and tasks assigned to them (Giangreco et al., 2013). Teachers have responsibility for leading instructional planning and decision-making, in addition to supervision and guidance of paraprofessionals. Paraprofessionals' tasks often vary based on the context in which they work, students they support, and supervisors' prior training and expectations. For example, paraprofessionals provide instructional and behavior support, either individually, in small groups, or sometimes class-wide, to maximize learning for students with greater behavioral or academic learning difficulties. In some instances, under the direction of teachers, paraprofessionals provide accommodations for academics (e.g., Riggs & Mueller, 2001) or social behavior (e.g., Fisher & Pleasants, 2011) to promote students' access to learning activities. In other instances, paraprofessionals deliver supplemental instruction or targeted interventions to individual students, or even small groups of students (e.g., Fuchs & Fuchs, 2007; Lane et al., 2007; Vadasy et al., 2006a, 2006b). Employed effectively, behavior interventions and accommodations can lead to reductions in classroom disruptive behavior, increases in on-task behavior, and improved learning (Penno et al., 2000). As with practices that impact student classroom behavior directly, instructional strategies used by paraprofessionals can promote greater student academic success by reducing barriers to participation (e.g., task difficulty) and increasing participation and learning (e.g., Gickling & Armstrong, 1978; Lee et al., 1999; Umbreit et al., 2004).

Federal law such as Every Student Succeeds Act of (2015) requires supervision, and encourages schools to allocate funds for ongoing training and supervision for paraprofessionals, yet gives minimal guidance on the type and delivery of professional development resources (e.g., Ashbaker & Morgan, 2006; Sobeck, 2016). For some time, special education scholars have raised serious concerns about the extent to which paraprofessionals are supervised, trained, and utilized in schools (e.g., Broer et al., 2005; Giangreco et al., 2013). Failure to meet paraprofessional needs for training and supervision has serious implications for the students they serve.

Based on a review of available research (39 studies), Farrell et al. (2010) described evidence of a trend in which paraprofessionals without adequate supervision, training, and performance feedback tended to produce limited student outcomes. For example, in a large-scale observational survey in the United Kingdom including

approximately 8000 first through eighth grade students and over 600 paraprofessionals, results indicated paraprofessionals received low levels of training and supervision across schools and grade levels, and were observed inconsistently using effective support strategies with students (Blatchford et al., 2009). Findings revealed that paraprofessionals often provided unclear or inaccurate explanations to students, engaged in limited or no monitoring of student understanding, and offered low-quality feedback to students. Such practices were associated with low gains in English/language arts and mathematics (controlling for prior achievement, special education status, and demographic factors).

Those who supervise and support the work of paraprofessionals—including special or general education teachers—report having few resources to assess, guide, and enhance the work of paraprofessionals in classrooms (e.g., Douglass et al., 2016; French, 1998; 2001; Reddy, Alperin, & Glover, 2020; Reddy, Lekwa, & Glover, 2020; Wallace et al., 2001). Although assessment practices currently in use to guide paraprofessional improvement remain unknown, research suggests that paraprofessionals receive limited training and ongoing support in classrooms (e.g., Broer et al., 2005; Giangreco et al., 2013; Reddy, Alperin, & Glover, 2020; Reddy, Lekwa, & Glover, 2020).

2 | PARAPROFESSIONAL IMPROVEMENT THROUGH ASSESSMENT

Researchers have found that supervisors of paraprofessionals (i.e., classroom teachers) need evidence-based approaches to objectively measure paraprofessional skills and deliver targeted performance feedback to improve paraprofessional skills (e.g., Carter et al., 2009; Drecktrah, 2000; French, 2001; Ghere & York-Barr, 2007; Wallace et al., 2001). Scholars agree that for training and supervision to be effective, supervisors of paraprofessionals must have reliable and valid assessment data on paraprofessional practices to generate specific feedback for improvement over time (Brock & Carter, 2013, 2017). Without sufficient professional development resources, such as valid assessments to identify needs for improvement in practices and generate data-driven performance feedback, paraprofessionals will likely not develop the skills necessary to effectively support classroom teachers and specifically meet the needs of students with or at risk for disabilities.

The present study is the first to critically examine the available assessments of paraprofessional practices and their psychometric properties for making score inferences of effectiveness or professional development. The purpose of this investigation was to address this gap in knowledge about assessment practices for these vital positions through two objectives. First, we conducted a systematic review to identify published and unpublished assessment approaches for paraprofessionals, followed by a descriptive review of assessment approaches developed by schools that were available publically online. Second, based on results from this review of assessment approaches, as well as the larger paraprofessional training literature, we outline recommendations for measurement development and validation including descriptions of possible constructs or domains of practices, and considerations for reliability and validity evidence aligned with purposes of evaluation and professional development. To this end, we addressed two research questions:

a. What is the nature and extent of available assessments of paraprofessional practices?
b. Do assessments have forms of reliability and validity of evidence?

3 | METHODS

3.1 | Assessment search and selection criteria

Information on instruments or procedures used to assess the practices of paraprofessionals in K-12 settings was gathered in two separate reviews: a systematic review of research literature and a comprehensive internet search

of paraprofessional assessment materials made available to the public. The search conducted in either the systematic review of research literature, or the review of publically available materials was not intended to generalize to the context of a specific set of paraprofessional practices, per se, but rather to describe trends *across* a population of paraprofessionals.

3.1.1 | Systematic review

Several methods were used to ensure a systematic review of the availability of paraprofessional assessments. First, a comprehensive search of research literature was conducted by using the keywords "paraprofessionals," "paraeducators," "teacher or classroom aides," "evaluations," "assessments," and "measures" within four databases from the years 1980 to 2019. The databases used were ERIC, Google Scholar, PsycInfo, and Sociological Abstracts. This yielded a total of 27 articles for initial review. Second, the references of these articles were checked, which did not identify any additional articles for review. Third, three distinguished scholars with expertise in paraprofessional training and supervision (i.e., Nancy French, Former Executive Director, Paraprofessional Resource and Research Center; Michael Giangreco, scholar of paraprofessional supervision and policy; Eric Carter, scholar in special education and issues surrounding professional development for paraprofessionals) were contacted to assist the authors in identifying published or unpublished studies (e.g., pending or completed dissertations) of paraprofessional assessments.

To be included in the review, instruments or assessment procedures for paraprofessional practices needed to have appeared in peer-reviewed articles, published research reports, or dissertations. In addition, identified instruments or procedures needed to be based on the context of paraprofessionals in K-12 educational settings, and had to include information about instrument or procedure design (e.g., constructs assessed, or method of assessment), or information about psychometric characteristics of the data produced (e.g., indices of reliability or validity). We excluded from review any examples of instruments or procedures that were (1) created to survey paraprofessional or teacher beliefs about issues such as paraprofessionals' training needs or views of support received, or (2) designed to assess the fidelity with which paraprofessionals implemented procedures or strategies associated with a specific intervention program. Also excluded from review were any instruments or assessment procedures focused on the practices of paraprofessionals from fields outside of education. A total of 3 instruments out of the 27 originally identified instruments met the criteria for review.

3.1.2 | Review of publically available instruments

We conducted an internet search subsequent to the systematic review of research literature to identify any publically available instruments or procedures for the assessment of paraprofessional practices. The same search terms and inclusion criteria used for the systematic review of research literature were applied (timeframe, K-12 context, and purpose of the instrument) with the exception that publically available tools need not to have been published in the research literature. A total of 11 articles or documents were identified and reviewed.

3.2 | Assessment review coding system

We applied a coding procedure based on a previously published systematic review (e.g., Reddy et al., 2009) to review the 14 identified instruments according to three dimensions: (a) instrument design and format; (b) presence and forms of validity evidence (specifically indices of reliability, forms of validity evidence, and theoretical foundation); and (c) constructs measured. Within the first dimension, we recorded the response format, number of

items or procedures, and scoring procedures. For the second dimension (evidence of validity) we determined whether theoretical foundations, or reliability or validity indices were reported for each instrument. This served as a quality indicator for each of the 14 identified instruments.

Within the third dimension (constructs assessed), we defined two broad categories based on prior research (e.g., Fisher & Pleasants, 2011; Giangreco et al., 2010), as well as credentialing and supervision standards for paraprofessionals (e.g., Council for Exceptional Children, 2015), under which the items or procedures from each instrument could be coded. These two broad categories were "Paraprofessional Interactions with Students" (any actions related to behavior management, accommodation, or instruction), and "Assessment" (any actions related to gathering or recording data on student behavior or performance). Items or procedures from each of the 14 identified instruments were coded under these broad categories in the first round of review. A substantial number of items did not represent interactions with students or assessment-related tasks, and two additional broad categories were added: "Paraprofessional Characteristics" (such as knowledge or physical characteristics), and "Professional" (including items about paraprofessionals' adherence to behavioral norms or expectations in the workplace, and relationships with other adults). We conducted a second round of review to add more specific constructs within the broad categories represented by the selected instruments. Finally, inter-rater agreement was obtained through a separate review by an advanced doctoral student using the full coding scheme; a rate of agreement of 92.45% was observed, and disagreements were resolved via discussion.

4 | RESULTS

As stated above, our exploration of research literature resulted in three instruments designed by researchers for the assessment of paraprofessional practices (or practices of individuals fulfilling roles typically occupied by paraprofessionals), and the subsequent internet search yielded a set of 11 assessment instruments created or used by school districts for purposes of paraprofessional evaluation, and made available online. Information about the design and scoring procedures of these 14 instruments, including an indication of the availability of psychometric characteristics is provided below and in Table 1 (instrument design and psychometrics). The instruments were predominantly in the format of rating scales ($n = 9$), but also included checklists ($n = 3$), one instrument with an open-ended, short response format, and one instrument that consisted of structured observation of paraprofessional interactions with students. The average number of items per instrument was approximately 27, yet substantial differences in length were observed, from a minimum of 4, up to a maximum of 62. Across all 14 instruments, a combination of theoretical foundation, reliability, and validity evidence was available for only 1 instrument; 3 additional instruments were based on an established body of theory, yet lacked any additional psychometric information. One instrument had published correlational validity evidence, without additional information about theoretical foundation or reliability. No psychometric information was available for the remaining nine instruments.

A total of 18 distinct constructs were targeted across measures based on the results of the initial and secondary reviews and inter-rater agreement check; operational definitions of each specific construct listed in Table 2. The extent to which each of the 14 instruments identified for this review targeted each of the 18 constructs is displayed in Table 3, which is a cross-tabulation of the number of items or procedures within each instrument that was coded under each of the 18 constructs.

The importance of a specific practice, as operationalized in instruments by researchers or district staff alike, might be judged by the frequency at which a construct appears in separate instruments. In this sample the constructs that appeared most often were elements of behavior management (i.e., antecedent and consequent strategies) and instructional delivery—each of these two construct categories was targeted in all 14 instruments we reviewed—followed by interactions with educators and relationships with students, which appeared in 11 of the instruments reviewed. Differences between constructs targeted by district-developed versus published

TABLE 1 Published and publically available assessments of paraprofessional practices

Measure	Origin	Format	# Items or Procedures	Scoring	Psychometric foundation
Anderson	LEA-Made	Rating scale	27	None	n
Bergenfield	LEA-Made	Rating scale	8	None	t
DISS observation	Research project	Structured observation	4	Sum	v
Doniphan	LEA-Made	Rating scale	28	None	n
Epping	LEA-Made	Rating scale	54	None	n
Everett	LEA-Made	Checklist	37	None	n
Instructional Assistant Task Assessment	Research Project	Checklist	62	Sum	n
Ionia	LEA-Made	Open-ended short response	21	None	n
North Hills	LEA-Made	Rating scale	25	None	n
Oak Park	LEA-Made	Checklist	39	None	n
QIDR	Research project	Rating scale	15	Sum	t, r, v
Utah Paras	LEA-Made	Rating scale	28	Sum	n
Wamego	LEA-Made	Rating scale	14	None	t
West New York	LEA-Made	Rating scale	18	Sum	t

Note: n = no psychometric foundation; r = reliability evidence available; t = instrument design was based on established theoretical framework; v = validity evidence available.

assessment approaches were noteworthy: among districts, items targeting paraprofessionals' workplace behaviors (such as confidentiality and professionalism) appeared frequently, but were not described at all in the three instruments we identified from the published research literature.

The relative importance of specific paraprofessional practice constructs might also be understood according to the number of items addressed to each. The most frequently coded construct categories for all items reviewed were "Instructional Delivery" and "Interaction with Educators" (78 and 73 times, respectively), followed by items focused on "Professionalism." Items focused on paraprofessional use of behavior management strategies (divided into antecedent strategies such as use of directives and proximity), were observed slightly more frequently than items describing paraprofessionals' assistance in preparation of "Instructional Materials", or use of accommodations or assistive technology to promote students' access to learning activities. Categories endorsed least frequently across all items were confidentiality, relationships with parents/families, and paraprofessionals' physical attributes.

4.1 | Published assessment approaches

The review yielded three published assessment approaches, described below in order of publication. The Instructional Assistant Task Assessment (IATA, Green & Barnes, 1988), is a semi-structured observation protocol offering a unique scope and metrics of paraprofessional practices. Specifically, the IATA was designed for gathering information about the work of paraprofessionals (such as provision of supervision to children during lunch or recess, or preparation of materials for lessons or classroom activities). Comprising binary checklist items

TABLE 2 Operational definitions of paraprofessional practice constructs

Construct	Operational definition
Paraprofessional characteristics	
Physical ability	Items describing physical requirements for tasks to be performed by a paraprofessional.
Academic skills & Ed. knowledge	Items concerning paraprofessionals' own literacy skills, numeracy, or skills in written expression, as well as their knowledge of education and learning.
Knowledge of student	A paraprofessional's awareness, knowledge, or understanding of a student or student(s) whom they serve.
Paraprofessional interactions with students	
Relationships with students	Items about how paraprofessionals establish and maintain rapport with students, including establishment of appropriate boundaries, expression of respect, positive attitudes, encouragement, and communicating norms of the classroom (social expectations). This did not include ratings of how paras use behavioral praise, or how paras encourage independence.
Student safety	Items in this category used the word "safety", or made direct references to paraprofessional tasks focused on self-care accommodation or intervention for students.
Accommodation & technology	Strategies used by paraprofessionals to remove barriers to student participation in learning or other school-based activities. This category includes items regarding assistive technology.
Student independence	Items that fit in this category used words such as "independence", or referred to paraprofessionals' appropriate provision of structure and support, or time and space for students to do something independently.
Behavioral antecedents	Antecedents included actions intended to obtain student attention or cue specific student behaviors. This included verbal or nonverbal directives, use of proximity (e.g., standing near a student to remind him/her to stay on task), or reviewing routines or classroom rules.
Behavioral consequences	Consequences include forms of behavioral feedback. Items in this category reflect things that paraprofessionals do *after* a student's behavior. Examples include verbal or nonverbal praise, punishment, or corrective feedback strategies.
Instructional materials	Items in this category refer to preparation of curricular materials, or other physical materials, for teaching. It also included items focused on planning or preparing for instruction. These items referred primarily to physical preparation for instruction.
Instructional delivery	This category referred to any act of teaching academic content. This involved direct communication (teaching) between a paraprofessional and students. It also referred to actions by the paraprofessional to reinforce instruction by a classroom teacher. Whereas the previous category focused on actions that preceded instruction, these items concerned the carrying out of instruction.
Assessment	
Assess behavior	Any item that talks about collecting data on student behavior. These used terms such as "watching" or "monitoring" student behavior, or specified more formal tasks such as tracking student behavioral data on record sheets.
Assess academics	This referred to any data collected by paraprofessionals about students' academic skills. These involved any kind of testing (i.e., worksheets or other forms of academic assessment).
Professional	
Clerical	These items reflected paperwork, or other office tasks, that were not related to any of the other previous categories.

(Continues)

TABLE 2 (Continued)

Construct	Operational definition
Confidentiality	Items in this category used the word "confidentiality" directly in the item. These were concerned with the degree to which paraprofessionals maintained or protected confidentiality for the students they served.
Interaction with educators	These items focused on what the para does *with* or *for* other adults in the school. There were two types of interactions: (1) collaboration or helping, or (2) compliance (such as following directions, or attendance at school). These items considered interactions between paraprofessionals and professional educators and administrators.
Professionalism	Elements of personal demeanor or behaviors that are expected in the workplace. This included work ethic, professional appearance, management of time and stress, demonstration of flexibility, staying on task, thinking ahead, or self-reflection and personal growth. This did not include items related to attendance at school.
Rapport with parents	These items referred specifically to paraprofessionals' relationships and interactions with students' parents.

(e.g., "completed" and "not completed"), the IATA was scored by taking the sum of all tasks completed by the paraprofessional. While the IATA appears to be the first published assessment developed to assess the work of paraprofessionals, the instrument focused broadly on paraprofessionals' task completion rather than specific skills implemented with children, and did not undergo any research to establish evidence of reliability, validity, and utility for professional development.

As part of the DISS project in the United Kingdom, Blatchford et al. (2009) adopted an inductive qualitative assessment approach that assessed paraprofessionals' and students' interactions. This procedure involved the pairing of quantitative information on time spent in various interactions (e.g., moving around the classroom, working with an individual student, and working with a group of students) with qualitative notes on the paraprofessional's interactions with students and the teacher's role during the lesson. The approach was intended to collect information that would enhance knowledge of how paraprofessionals interact with students and teachers, and how this interaction is associated with students' academic and behavioral functioning. Based on a sample of paraprofessionals, teachers, and students in years 1, 3, and 7, and 10 of schooling in 21 separate schools in the United Kingdom, Blatchford et al. (2009) found the four constructs defined in their observation protocol to have generally significant, negative associations between increasing degrees of paraprofessional involvement with students and students' levels of academic achievement by the end of the year. Although two positive relationships were observed between paraprofessional presence and student achievement in science in primary grades (between about 20% and 100% of a "national curriculum level;" see Sainsbury & Sizmur, 1998, for detail about this metric), the majority of measured student outcomes demonstrated negative relationships with measured paraprofessional constructs, ranging between −20% and −85% of a national curriculum level. No information about the reliability of the data from this procedure was available. Although the data generated from this approach allowed researchers to examine the association between paraprofessional behavior and student achievement, it has limited psychometric evidence beyond the association with student academic and behavioral functioning.

Harn et al. (2012) created the Quality of Intervention Delivery and Receipt (QIDR) rating scale based on the theoretical framework of Explicit Instruction (e.g., Archer & Hughes, 2011). The QIDR was designed to be a nonsubject specific assessment of intervention implementation practices that consists of 15 items that describe a set of instructional and behavior management processes for interventionists. Each item may receive a score of 0 ("no implementation"), 1 ("inconsistent implementation"), 2 ("effective implementation"), or 3 ("expert implementation"). At this time, information about the psychometric qualities of the QIDR includes preliminary

TABLE 3 Assessment contents by paraprofessional practice construct

		Paraprofessional characteristics			Paraprofessional interactions with students								Assessment		Professional					
Measure	Items or procedures[a]	Physical ability	Academic skills/ed. knowledge	Knowledge of student	Relationship w/students	Student safety	Accommodation tech.	Student independence	Behavioral antecedents	Behavioral consequences	Instructional materials	Instructional delivery	Assess behavior	Assess academics	Clerical	Confidentiality	Interaction w/educators	Professionalism	Rapport w/parents	N/A
Anderson	27		1	3	1			1	3	2		2			1	1	11	4		
Bergenfield	8				3				1	1		5		1			1			
DISS observation[b]	4				1				2	1		1								
Doniphan	28		1	5	2			1	1	1		1			1	1	12	3		
Epping	54	1	1	2	4	3	1	3	1	1	2	5	1	2	1	1	20	8	1	
Everett	37		3		1	3	2		3	1		6	1	1	2	2	10	7	1	1
Instr. assistant task assessment	62					3	2		6	5	11	23		5	7		1		3	2
Ionia	21						11	3	3	3		2								2
North Hills	25				1				2	3	2	4	6	5	2		3	2		
Oak Park	39		1		5	1	1	2	3	3	3	7		1	1	1	7	7	2	1
QIDR	15				1				2	2	2	8								
Utah Paras	28								2	2	4	10	1	1	1	1	4	3		1

(Continues)

TABLE 3 (Continued)

Measure	Items or procedures[a]	Paraprofessional characteristics			Paraprofessional interactions with students								Assessment		Professional					
		Physical ability	Academic skills/ed. knowledge	Knowledge of student	Relationship w/students	Student safety	Accommodation tech.	Student independence	Behavioral antecedents	Behavioral consequences	Instructional materials	Instructional delivery	Assess behavior	Assess academics	Clerical	Confidentiality	Interaction w/educators	Professionalism	Rapport w/parents	N/A
Wamego	14		1	1	2	1		1	1	1	1	1				1	2	3		
West New York	18		2		2		1	1	1	1		3	1	1			2	3		2
Total		*1*	*10*	*11*	*23*	*11*	*18*	*12*	*31*	*27*	*25*	*78*	*10*	*17*	*16*	*8*	*73*	*40*	*7*	*9*

[a]Items or procedures could be coded under multiple categories; the sum of codes can be greater than the total number of items or procedures within an instrument.

[b]This instrument was developed in the United Kingdom; all other instruments identified for this review were developed in the United States.

documentation of adequate reliability and validity based on a sample of QIDR data from a group of 11 instructional assistants (paraprofessionals) implementing supplemental reading interventions under the supervision of licensed teachers to 31 children in kindergarten in the northwestern United States. Spear (2015) analyzed relationships between QIDR, the CLASS (a similar measure, designed to target teacher–student interactions; Pianta et al., 2008), as well as relationships between QIDR data and students' development of basic reading skills. Results indicated evidence of strong concurrent relationships between QIDR and CLASS (between 0.80 and 0.90), and a predictive relationship in which QIDR data accounted for approximately 36% of the group-level variability in student reading outcomes. Reliability was examined in a subsequent analysis of the same data set by Fritz et al. (2019): results included an intra-class correlation of 0.71, indicating strong reliability in observations as short in duration as 10 min.

4.2 | District-adapted or developed assessment approaches

A set of 11 district-developed assessment approaches were identified in this review; each varied in scope and method for assessing paraprofessional practices. Several of these approaches—"Bergenfield," "Wamego," and "West New York"—were adaptations of the Danielson Framework for Teaching (FFT; Danielson, 2013) one of the most widely used teacher evaluation approaches. The FFT, a standards-based observational framework assesses qualities of teachers' practices in four broad domains (i.e., Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities) that include 22 components of practices that are grounded in constructivist models of teaching. Adaptations of FFT for the purpose of assessing paraprofessional practices focused primarily upon FFT Domains 2 (Classroom Environment) and 3 (Instruction). Exemplars of practices within those domains were then altered by district staff to reflect the different roles and expectations for paraprofessionals. Modifications of the FFT in these instruments generally included (a) specification of "paraprofessional" in operational definitions of practices, (b) a focus on the paraprofessional's role as an assistant, rather than leader, in planning and carrying out instruction, and (c) additions of language reflecting roles in provision of accommodation, and interactions or collaboration with supervising teachers.

In other cases, school districts have generated their own instruments by defining specific domains of practice by paraprofessionals informed in part by professional and local standards. These instruments were primarily in the form of rating scales ($n = 8$) used for purposes of a summative evaluation of paraprofessionals; two instruments comprised checklists (similar to the IATA, above) and the 11th instrument consisted of open-ended questions to identify areas of strength and needs for development in paraprofessionals. This subgroup of 11 instruments comprised an average of 27 items each, and used sums or ratios in scoring. Whereas the researcher-developed instruments described above tended to focus on aspects of instruction and behavior management for students, the district-developed instruments included a more broad focus, and included items targeting aspects of paraprofessionals' knowledge of students and educational principles, and consistently addressed items to elements of professionalism (i.e., confidentiality, punctuality, and professional appearance and demeanors). The psychometric foundations of the district-developed instruments were limited to the use of established theories or frameworks (e.g., FFT); no evidence of examination of reliability or validity evidence was available.

5 | DISCUSSION

The relationship between formative assessment or evaluation of teacher practices and enhancement of instruction via professional development has been established in educational research (e.g., Reddy et al., 2017, 2009; Reddy, Alperin, & Glover, 2020; Reddy, Lekwa, & Glover, 2020). This knowledge may serve as an important example for research concerning paraprofessionals. We assert that accurate measurement of paraprofessional practices could

be critical to providing performance feedback that can be used to guide professional development conversations and supports. This assessment review indicates that instrumentation for assessing paraprofessionals' practices with students and classroom teachers is extremely scarce—a significant gap in resources for assessing and improving paraprofessional competencies in supporting students with or at risk for special education services. Research and development on key domains of practice likely has been challenged by a lack of clarity around the roles of paraprofessionals, plus limited knowledge of the impact of these roles on students' growth and functioning in schools. Findings from this present review of available assessments for paraprofessionals serve as a first step in bridging this study to practice gap by indicating (a) what aspects of paraprofessional job performance are most important to schools and (b) how schools operationalize their expectations for paraprofessionals' practices in support of students in need for special education services.

5.1 | A proposed framework for development and validation of paraprofessional assessments

Based on information from this review, and drawing on the effective teaching and paraprofessional literatures, we propose a framework for measurement development and validation efforts to springboard innovations in this under-studied area. First, we describe a set of possible constructs for measurement development that are reflective of the results described above, but also informed by the literature underlying practices highlighted from instruments reviewed in this study. Second, we outline recommendations for programs of research to follow in the establishment of psychometric evidence, including traditional elements of reliability and validity, as well as additional validity considerations including assessment utility or feasibility by school personnel, sensitivity to change in scores (practices) following professional development, and predictive relationships of scores reflecting qualities of paraprofessional practices to student outcomes.

5.1.1 | Paraprofessional practice constructs

Although decades of research on effective instruction and behavior management provide a nomenclature for describing how adult behaviors can function to promote student attention, participation, and learning (e.g., Brophy & Good, 1986; Burns et al., 2008; Rosenshine & Stevens, 1986; Ysseldyke & Christenson, 1993), this study is typically focused on, and was written for, professional educators rather than paraprofessionals. Thus, the first step in the development of assessments of paraprofessional practices is the operationalization of constructs reflecting practices that (a) align with schools' and students' needs, and (b) are supported by established theory and empirical evidence.

Drawing first on the information gained in our current review of current assessment approaches, we focus our framework on a variety of strategies that have been described and recommended for use in the evaluation of the supplemental instructional resources received by students with or at risk for requiring special education services (e.g., Gersten et al., 2005; Kretlow & Bartholomew, 2010; Mowbray et al., 2003; Pianta & Hamre, 2009; Reinke et al., 2016). Although the scope of paraprofessionals' roles in the classroom is limited, numerous organizations have focused on paraprofessionals' knowledge and use of such strategies with students (American Federation of Teachers, 1998; Council for Exceptional Children, 2015). We propose a set of constructs that is sufficiently broad to align with a variety of common paraprofessional roles, but also differentiated from elements of teaching that are, by legal requirement and best practice, reserved for certified teachers. These suggested constructs or domains (see Figure 1) may include practices that: (a) promote engagement and positive behavior, (b) support instruction, (c) provide effective feedback for students, and (4) enhance relationships with students. The constructs we propose, based on this review, focus primarily on ways in which paraprofessionals interact with students, as this

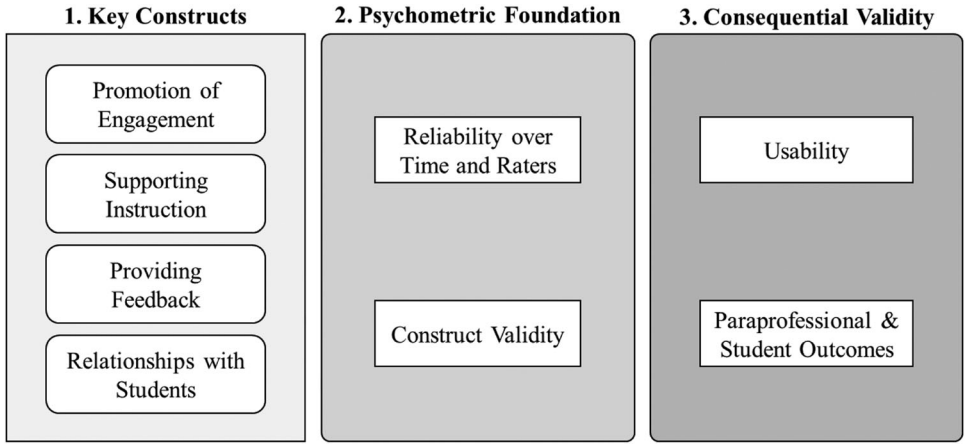

FIGURE 1 A framework for development and validation of paraprofessional assessments

domain would encompass the core sets of skills educators might aim to enhance through paraprofessional supervision and professional development. Other domains identified in this review, such as those reflecting paraprofessional relationships with educators, may offer productive avenues for future research.

Promotion of student engagement and positive behavior

Strategies paraprofessionals might use to help guide or improve the behavioral functioning of students they support were represented in each of the instruments identified for this review. Among the strategies included under "interactions with students", antecedent strategies to assist students with behavior received among the highest number of items or procedures (second only to strategies to support instruction). These strategies, intended to prompt student participation in learning activities, appear to represent one of a paraprofessional's key roles in supporting student learning. Any new measures of paraprofessional practices should therefore target antecedent behavioral strategies to be aligned with paraprofessional roles and needs for support.

For students to benefit from instruction they must be engaged during learning activities (e.g., participate in learning activities) and regulate their behavior (e.g., attend to instructional stimuli, complete tasks) during academic tasks and/or interactions (e.g., Gettinger & Seibert, 2002; Reddy et al., 2018; Singh et al., 2002; Wentzel, 1993). Examples of strategies useful in the promotion of student engagement and behavior include directives, careful use of physical proximity. Directives are verbal statements used to communicate to students exactly what behavior is expected, or as a cue for the student to engage in the desired behavior. Directives have the greatest effect when they are short, describe specific observable behaviors, issued one at a time, and followed by a strategic pause (e.g., Matheson & Shriver, 2005; Roberts et al., 1978). The physical presence of the adult near the student can also be used to help promote and maintain students' engagement with tasks; for paraprofessionals, general proximity of 2 ft (close proximity), or within 6 ft (medium proximity) has been found to be effective in reducing the frequency of inappropriate or off-task behaviors in the general education classroom (Werts et al., 2001). Additionally, there is a substantial research base on the effectiveness of classroom behavior interventions for enhancing student engagement and social behaviors in schools (e.g., DuPaul et al., 2012). Collectively, this literature highlights a range of practices such as function-based interventions (e.g., acquisition, attention seeking/reinforcement, escape, and generalization) that could be used by paraprofessionals.

Depending on the nature of the learning activity and students' levels of skill or ability and social functioning, paraprofessionals must be able to judge the degree to which students would benefit from more, or less direct involvement of the paraprofessional by responding to student requests for assistance or preventing frustration and disengagement. For example, students presented with tasks they cannot yet complete accurately might be more

likely to engage in off-task or inappropriate behaviors and require a greater degree of modeling (Szadokierski et al., 2017), practice, and feedback on the accuracy of their responses (Eckert et al., 2006). Similarly, by monitoring the way in which students attend to and participate in learning activities, paraprofessionals can judge the need for antecedent strategies to promote behavior or engagement such as use or nonuse of physical proximity (e.g., Feldman & Matos, 2013; Malmgren & Causton-Theoharis, 2006) and pre-correction and clear directives (e.g., Kern & Clemens, 2007).

Support for instruction

As with strategies to promote students' behavioral functioning, instructional strategies appeared to be another essential domain of paraprofessional practice. Strategies related to working with instructional materials or providing some form of instruction to students appeared in each of the instruments identified in this review, and accounted for the single largest proportion of instrument content.

Whether working to provide accommodations in an inclusive setting for children with disabilities, or providing supplemental academic or social skill interventions, paraprofessionals often use instructional strategies to supplement or reinforce previously taught, but unmastered skills. Students with disabilities that impact academic learning, or students with related skill deficits, often benefit substantially from instruction that is explicit, incorporating clear explanation, and sufficient modeling (e.g., Kroesbergen & Van Luit, 2003; Foorman & Torgeson, 2001). The degree to which students comprehend and later recall information or skills depends in part on the amount and quality of instructional presentation they receive. Examples might include the use of clear definitions during supplemental vocabulary activities (e.g., Stahl & Fairbanks, 1986), summarization and demonstration (e.g., Harbour et al., 2015; Haston, 2007), or more involved modeling strategies such as hand-over-hand demonstrations for nonimitative students (e.g., Carr et al., 1978; Carter et al., 2009).

Provision of feedback

As with the provision of behavioral antecedents, strategies involving specific consequences in response to student behavior appeared to some extent in each of the instruments included in this review. Qualities of the feedback provided by paraprofessionals can be useful for promoting student behavior and academic engagement in learning activities, increasing students' content knowledge, and reinforcing the accuracy of students' performance or skills. Feedback may be divided into two forms: praise and corrective feedback. Praise, as a strategy for shaping behavior and maintaining engagement during instruction, is a technique that is implemented less often than behavioral corrective feedback (e.g., Floress et al., 2017; Reddy et al., 2013; White, 1975). Praise that is specific and offered immediately is a useful method for reinforcing the occurrence of desired behaviors in students (Chalk & Bizo, 2004; Henderlong & Lepper, 2002). Corrective feedback for inappropriate social behavior or inaccurate academic performance is verbal or nonverbal statements or gestures provided by teachers to redirect inappropriate behavior. Like behavior or academic praise, corrective feedback should be specifically labeled and given immediately after inappropriate behavior is observed (Bangert-Drowns et al., 1991). Corrective feedback serves as an important instructional tool when paraprofessionals describe specifically what was incorrect about the student's behavior or academic response *and* what appropriate behavior or response should be used going forward.

Development of positive relationships with students

The qualities of relationships between paraprofessionals and students also appeared consistently throughout the instruments included in this review. All but three of the instruments reviewed included procedures or items assessing the degree to which paraprofessionals establish and maintain good interpersonal relationships with the students they support.

How paraprofessionals forge positive relationships with the students they serve can influence students' engagement, learning, social behavior, and independence. Students who report positive connections with school personnel exhibit a greater sense of school belongingness, connection, climate, and overall well-being

(e.g., Baker, 1999; McGrath & Van Bergen, 2015; Wentzel, 2010). Specifically, the degree to which students feel liked, supported, encouraged, and empowered by the adults that support them in school has an important effect on students' emotional, behavioral, and academic outcomes (e.g., Hamre & Pianta, 2006; Roorda et al., 2011). Paraprofessionals can enhance relationships with their students with a variety of strategies such as active listening, use of nonverbal encouragement (such as smiles, nods, signals), and/or validating students' interests or concerns (e.g., Driscoll & Pianta, 2010; Murray & Malmgren, 2005).

5.1.2 | Paraprofessional measurement validation

Similar to other evaluation approaches for school personnel (e.g., teacher observational assessment), the value of paraprofessional assessments can be captured through a comprehensive program of research examining the consistency of scores (reliability), inferences that can be drawn from sets of scores (construct validity), and consequences of using sets of scores in routine educational practices (external validity). Furthermore, the establishment of reliability and validity evidence must consider the intended purpose, user(s), and context (i.e., general education vs. special education contexts) in which new tools will be utilized. Below, we offer recommendations tailored to the development and validation of instruments for assessing the practices of paraprofessionals.

Considerations for reliability evidence

In addition to traditional indices of reliability of scores from single points in time, assessments of paraprofessional practices should examine forms of reliability evidence such as the stability of ratings over time (enabling detection of changing trends in paraprofessional practices), and consistency of independent supervisor ratings of paraprofessional practices during the same classroom context or point in time (i.e., interobserver reliability). While all indices of reliability are important, tools that yield similar scores across trained paraprofessional supervisors are critical for ensuring school personnel confidence that score inferences are informative for training or supervision for individuals, plus fair and balanced for human capital decision-making (e.g., paraprofessional retention, promotion, and professional improvement plans). Inter-rater reliability is particularly salient for making personnel decisions for individual versus school-wide decisions for these support positions.

Considerations for construct validity evidence

Subsequent to the definition of paraprofessional practice constructs, a program of paraprofessional measurement development must address several issues, including evaluation of instrument content (common evidence-based practices), response processes, and relations of constructs to other variables and internal structure (American Educational Research Association American Psychological Association & National Council on Measurement in Education, 2014). Within the present context, evidence on content validity would be the degree to which content of a paraprofessional assessment appears to represent important and common practices that were designed to measure. Thus, the content of a paraprofessional tool should be drawn from research-based instructional and classroom management practices, as well as current national and/or state credentialing standards. A common method for testing content validity is to have supervisors of paraprofessionals and/or experts of paraprofessional training and supervision rate the importance of specific practices (items) for measuring the intended constructs. Additionally, data from surveys of paraprofessionals' roles and practices, or from surveys of schools' practices in the assessment and evaluation of paraprofessionals may be helpful in establishing valid measurement content.

An important, but often overlooked form of validity evidence is response processes. Response process refers to whether paraprofessional supervisors and paraprofessionals of a tool respond to or interpret it in ways as intended by the developers (e.g., refer to strategies described by Myford & Wolfe, 2003, 2004). While this form of validity may be challenging to ascertain, it does offer valuable insight into the commonly observed classroom

practices of paraprofessionals and how intended users (supervisors and paraprofessionals) understand and use these practices. Such evidence might be gathered in a variety of ways, including analyses of ratings collected from multiple informants for one paraprofessional (i.e., self- and supervisor ratings), or also analyses of correspondence between ratings of video-recorded interactions between paraprofessionals, students, and educators.

The validation of paraprofessional assessments should examine the degree to which scores of the tool converge with scores from measures of similar constructs (e.g., FFT behavior management domain) and diverge from scores of unrelated constructs (e.g., FFT preparation and planning for leading instruction). While evidence based on relations to other variables may be challenging to obtain due to the limited number of existing tools for paraprofessionals, this form of validity is warranted to ensure valid representation.

Given the complex range of potential interactions between paraprofessionals, students, and teachers, development of assessments of paraprofessional practices likely requires the use of rating scales versus a reliance on direct observation and counts of discretely observable behaviors. As previously noted, construct validity can be addressed through testing how scales and items of a paraprofessional assessment fit together as intended. Testing the internal structure of new paraprofessional tools can be examined through methods such as point-biserial correlations, factor analytic methods (exploratory and/or confirmatory analyses), as well as item response theory approaches (e.g., differential item functioning; Steinberg & Thissen, 2006; Zumbo, 1999). Results of such analyses may be useful to researchers and measurement developers throughout the measurement design and validation process, as efforts to improve the fit of scales and items with theoretically important constructs should lead to instruments with greater sensitivity to intra and interindividual differences in the practices of paraprofessionals.

Considerations for consequential validity evidence

This form of validity refers to the ability of new tools to relate to and predict important paraprofessional outcomes (e.g., retention, promotion, professional improvement plans) and student academic and behavior outcomes. While consequential validity is essential for all assessments, we recommend that development and validation of new paraprofessional tools gather evidence on school personnel's perceived usability, sensitivity to change following professional development (e.g., coaching intervention), and predictiveness of key paraprofessional and student outcomes.

Usability

The evaluation of the usability of new tools to assess paraprofessional practices should include perceived acceptability of the tool (i.e., purpose, design, scales, or items) from both supervisors of paraprofessionals and paraprofessionals themselves. Factors to guide the testing of usability may also include the perceived time and costs of administration, feasibility of use, clarity of instructions, management of data for formative and summative evaluation purposes, and linkages of scores to professional improvement planning.

Related to school personnel's perceptions of the feasibility and usefulness of a new instrument is the scalability of the use of that new instrument. We advocate for assessment systems that are multimethod and multisource by design; such systems can easily become a logistic burden. By incorporating technology to facilitate these assessment processes, from data collection to scoring, data storage, and reporting, logistic barriers to full implementation of a multicomponent assessment system such as we propose can be reduced. This involves consideration of technology at the initial stages of measurement design to facilitate scaling up of the instrument to digital platforms. Making use of digital platforms for assessment will not only help reduce logistic burdens in data collection and use, but may also increase data visualization capabilities. The format and ways in which data are communicated to school personnel can be critical for the ultimate quality of data uses or decisions made for enhancing practices. For example, high-quality visualizations may be used to convey information in a format more understandable than that of technical information such as *Z*-scores or confidence intervals.

Relationship to desirable outcomes

The scores that are generated by instruments designed to promote improvement in paraprofessional practices must demonstrate two important properties: (a) relationships with outcomes that are connected with the measured constructs (i.e., domains of practice), and (b) sensitivity to change in constructs in response to training or intervention for individuals (i.e., paraprofessionals) measured repeatedly over time by supervisors. Although evidence of predictive validity is often established through analysis of measurements that are quite distal in time, for the purpose of assessing and improving paraprofessional practices, we advocate for evidence of predictive validity for more immediate student outcomes. For example, such evidence might involve the prediction of student outcomes at the conclusion of a lesson, later in a school day, or across a series of weeks. The majority of instruments identified for this review prioritized assessment of paraprofessionals' support for student behavior, followed by an assessment of a variety of ways in which paraprofessionals help teachers in the delivery of instruction. Measures of paraprofessional practices in either domain should demonstrate a predictive relationship with student academic and behavioral functioning in classrooms over time. Models analyzing such a relationship should be able to demonstrate higher levels of student functioning controlling for disability status and prior skill levels after working with a paraprofessional (e.g., Blatchford et al., 2009).

Finally, an important consideration for establishing the reliability of new paraprofessional assessments is that they yield reliable estimates of change or growth in paraprofessional practices. Similar to student formative assessment, teacher or paraprofessional assessment approaches need to reliably capture domains of practice (i.e., strategies that enhance student engagement and positive behavior) across multiple measurements. The degree to which domain scores obtained from repeated measures over time reflect actual rates of change in paraprofessional practices is an important prerequisite for utility in the training and supervision of paraprofessional improvement efforts. Such repeated assessment data should not only help identify when specific practices require improvement, but should reliably indicate the precision and magnitudes of changes in classroom skills over time (e.g., Fuchs, 2004; Lekwa et al., 2020).

6 | CONCLUSION

At this time schools' employment and reliance on paraprofessionals for the support of students at risk, or with disabilities has outstretched research and practice surrounding the effective supervision and guidance of individuals in this workforce. Teachers have very limited supervision training and professional development resources for effectively supporting paraprofessional instructional support skills and implementation of best practices that meet the continuum of students' needs in the classroom. In this article, we articulated an urgent need for measurement development and validation for paraprofessionals. To this end, we offer recommendations to stimulate future research and innovation in measurement development and validation for this large, and under-supported staff in schools.

ACKNOWLEDGMENT

The research reported here was supported by the Institute of Education Sciences at Rutgers University (Grant no. R324A170069). The opinions expressed are those of the authors and do not represent views of the IES.

ORCID

Adam J. Lekwa http://orcid.org/0000-0002-6870-7948
Linda A. Reddy http://orcid.org/0000-0001-8314-2810

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

American Federation of Teachers. (1998). *Standards for a profession*.

Archer, A., & Hughes, C. A. (2011). *Explicit instruction: Efficient and effective teaching*. Guilford Publications.

Ashbaker, B. Y., & Morgan, J. (2006). The role of administrators in paraprofessional supervision to support ethnic minority students with special needs. *Educational Considerations*, *34*, 17–31.

Baker, J. A. (1999). Teacher-student interaction in urban at-risk classrooms: Differential behavior, relationship quality, and student satisfaction with school. *The Elementary School Journal*, *100*, 57–70.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*, 213–238.

Blalock, G. (1991). Paraprofessionals: Critical team members in our special education programs-how to successfully utilize the paraprofessional in the special education setting. *Intervention in School and Clinic*, *26*, 200–215.

Blatchford, P., Bassett, P., Brown, P., Martin, C., Russell, A., & Webster, R. (2009). *Deployment and impact of support staff project. Research summary. Short summary of the main findings, conclusions and recommendations from the DISS project*.

Brock, M. E., & Carter, E. W. (2013). A systematic review of paraprofessional-delivered educational practices to improve outcomes for students with intellectual and developmental disabilities. *Research and Practice for Persons with Severe Disabilities*, *38*, 211–221. https://doi.org/10.1177/154079691303800401

Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education*, *38*, 131–144. https://doi.org/10.1177/0741932516653477

Broer, S. M., Doyle, M. B., & Giangreco, M. F. (2005). Perspectives of students with intellectual disabilities about their experiences with paraprofessional support. *Exceptional Children*, *71*, 415–430.

Brophy, J. E., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research in teaching* (3rd ed., pp. 328–375). Macmillan.

Burns, M. K., VanDerHeyden, A. M., & Boice, C. H. (2008). Best practices in delivery of intensive academic interventions. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 1151–1162). National Association of School Psychologists.

Carr, E. G., Binkoff, J. A., Kologinsky, E., & Eddy, M. (1978). Acquisition of sign language by autistic children. I: Expressive labelling. *Journal of Applied Behavior Analysis*, *11*, 489–501.

Carter, E., O'Rourke, L., Sisco, L. G., & Pelsue, D. (2009). Knowledge, responsibilities, and training needs of paraprofessionals in elementary and secondary schools. *Remedial and Special Education*, *30*, 344–359. https://doi.org/10.1177/0741932508324399

Chalk, K., & Bizo, L. A. (2004). Specific praise improves on-task behaviour and numeracy enjoyment: A study of year four pupils engaged in the numeracy hour. *Educational Psychology in Practice*, *20*, 335–351.

Council for Exceptional Children. (2015). *What every special educator must know: Professional ethics and standards*.

Danielson, C. (2013). *The framework for teaching evaluation instrument, 2013 instructionally focused edition*. The Danielson Group. http://www.danielsongroup.org/framework

Douglass, S. N., Chapin, S. E., & Nolan, J. F. (2016). Special education teachers' experiences supporting and supervising paraeducators: Implications for special and general education settings. *Teacher Education and Special Education*, *39*, 60–74. https://doi.org/10.1177/0888406415616443

Drecktrah, M. E. (2000). Preservice teachers' preparation to work with paraeducators. *Teacher Education and Special Education*, *23*, 157–164.

Driscoll, K. C., & Pianta, R. C. (2010). Banking time in head start: Early efficacy of an intervention designed to promote supportive teacher–child relationships. *Early Education and Development*, *21*, 38–64.

Dulfer, K. J. (2013). *The design, implementation, and formative evaluation of a classroom aide professional development training program* (Doctoral dissertation). Rutgers University-Graduate School of Applied and Professional Psychology.

DuPaul, G. J., Eckert, T. L., & Vilardo, B. (2012). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis 1996-2010. *School Psychology Review*, *41*, 387–412.

Eckert, T. L., Dunn, E. K., & Ardoin, S. P. (2006). The effects of alternate forms of performance feedback on elementary-aged students' oral reading fluency. *Journal of Behavioral Education*, *15*, 148–161.

Every Student Succeeds Act of 2015 (ESSA), 20 U.S.C § 6311. *et seq*. (2015).

Farrell, P., Alborz, A., Howes, A., & Pearson, D. (2010). The impact of teaching assistants on improving pupils' academic achievement in mainstream schools: A review of the literature. *Educational Review*, *62*, 435–448. https://doi.org/10.1080/00131911.2010.486476

Feldman, E. K., & Matos, R. (2013). Training paraprofessionals to facilitate social interactions between children with autism and their typically developing peers. *Journal of Positive Behavior Interventions*, *15*, 169–179. https://doi.org/10.1177/1098300712457421

Fisher, M., & Pleasants, S. L. (2011). Roles, responsibilities, and concerns of paraeducators: Findings from a statewide survey. *Remedial and Special Education*, *33*, 287–297.

Floress, M. T., Jenkins, L. N., Reinke, W. M., & McKown, L. (2017). General education teachers' natural rates of praise: A preliminary investigation. *Behavioral Disorders*, *43*, 411–422. https://doi.org/10.1177/0198742917709472

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. Learning Disabilities. *Research & Practice*, *16*, 203–212.

French, N. K. (1998). Working together: Resource teachers and paraeducators. *Remedial and Special Education*, *19*, 357–368.

French, N. K. (2001). Supervising paraprofessionals: A survey of teacher practices. *The Journal of Special Education*, *35*, 41–53.

Fritz, R., Harn, B., Biancarosa, G., Lucero, A., & Flannery, K. B. (2019). How much is enough? Evaluating intervention implementation efficiently. *Assessment for Effective Intervention*, *44*(2), 135–144. https://doi.org/10.1177/1534508418772

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–193.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, *39*, 14–20.

Gersten, R., Fuchs, L., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*, 149–164.

Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (Vol. IV, pp. 773–787). The National Association of School Psychologists.

Ghere, G., & York-Barr, J. (2007). Paraprofessional turnover and retention in inclusive programs: Hidden costs and promising practices. *Remedial and Special Education*, *28*, 21–31.

Giangreco, M. F., Suter, J. C., & Doyle, M. B. (2010). Paraprofessionals in inclusive schools: A review of recent research. *Journal of educational and psychological consultation*, *20*, 41–57.

Giangreco, M. F., Suter, J. C., & Hurley, S. M. (2013). Revisiting personnel utilization in inclusion-oriented schools. *The Journal of Special Education*, *47*, 121–132.

Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities*, *11*, 559–566.

Green, J. E., & Barnes, D. L. (1988). Do your "aides" aid instruction? A tool for assessing the use of paraprofessionals as instructional assistants. *The Teacher Educator*, *24*, 2–9.

Hamre, B. K., & Pianta, R. C. (2006). Student-teacher relationships. In G. G. Bear, & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 59–71). National Association of School Psychologists.

Harbour, K. E., Evanovich, L. L., Sweigart, C. A., & Hughes, L. E. (2015). A brief review of effective teaching practices that maximize student engagement. *Preventing School Failure: Alternative Education for Children and Youth*, *59*, 5–13.

Harn, B. A., Forbes-Spear, C., Fritz, R., & Berg, T. (2012). *Quality of Intervention Delivery and Receipt (QIDR) observation tool*. University of Oregon.

Haston, W. (2007). Teacher modeling as an effective teaching strategy. *Music Educators Journal*, *93*, 26–30.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, *128*, 774–795. https://doi.org/10.1037/0033-2909.128.5.774

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools*, *44*, 65–75.

Kretlow, A. G., & Bartholomew, C. C. (2010). Using coaching to improve the fidelity of evidence-based practices: A review of studies. *Teacher Education and Special Education*, *33*, 279–299.

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education*, *24*, 97–114.

Lane, K. L., Fletcher, T., Carter, E. W., Dejud, C., & Delorenzo, J. (2007). Paraprofessional-led phonological awareness training with youngsters at risk for reading and behavioral concerns. *Remedial and Special Education*, *28*, 266–276.

Lee, Y.-Y., Sugai, G., & Horner, R. H. (1999). Using an instructional intervention to reduce problem and off-task behaviors. *Journal of Positive Behavior Interventions*, *1*, 195–204. https://doi.org/10.1177/109830079900100402

Lekwa, A. J., Reddy, L. A., & Shernoff, E. S. (2020). The magnitude and precision of estimates of change in formative teacher assessment. *School Psychology*, *35*, 137–145. https://doi.org/10.1037/spq0000355

Malmgren, K. W., & Causton-Theoharis, J. N. (2006). Boy in the bubble: Effects of paraprofessional proximity and other pedagogical decisions on the interactions of a student with behavioral disorders. *Journal of Research in Childhood Education*, *20*, 301–312. https://doi.org/10.1080/02568540609594569

Matheson, A. S., & Shriver, M. D. (2005). Training teachers to give effective commands: Effects on student compliance and academic behaviors. *School Psychology Review*, *34*, 202–219.

McGrath, K. F., & Van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student–teacher relationships and their outcomes. *Educational Research Review*, *14*, 1–17.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*, 315–340.

Murray, C., & Malmgren, K. (2005). Implementing a teacher–student relationship program in a high-poverty urban school: Effects on social, emotional, and academic adjustment and lessons learned. *Journal of School Psychology*, *43*, 137–152.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, *4*, 386–422.

Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, *5*, 189–227.

Penno, D. A., Frank, A. R., & Wacker, D. P. (2000). Instructional accommodations for adolescent students with severe emotional or behavioral disorders. *Behavioral Disorders*, *25*, 325–343.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*, 109–119.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system™: Manual K-3*. Paul H. Brookes Publishing.

Reddy, L. A., Alperin, A., & Glover, T. G. (2020). A critical review of professional development literature for paraprofessionals supporting students with externalizing behavior disorders. *Psychology in the Schools*, 1–22. https://doi.org/10.1002/pits.22381

Reddy, L. A., Cleary, T. J., Alperin, A., & Verdesco, A. (2018). A critical review of self-regulated learning interventions for children with attention-deficit hyperactivity disorder. *Psychology in the Schools*, *55*, 609–628.

Reddy, L. A., Dudek, C., & Lekwa, A. J. (2017). Classroom Strategies Coaching Model: Integration of formative assessment and instructional coaching. *Theory into Practice*, *56*, 46–55. https://doi.org/10.1080/00405841.2016.1241944

Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology*, *51*, 683–700.

Reddy, L. A., Lekwa, A. J., & Glover, T. A. (2020). Supporting paraprofessionals in schools: Current research and practice. *Psychology in the Schools*, https://doi.org/10.1002/pits.22457

Reddy, L. A., Newman, E., De Thomas, C. A., & Chun, V. (2009). Effectiveness of school-based prevention and intervention programs for children and adolescents with emotional disturbance: A meta-analysis. *Journal of School Psychology*, *47*, 77–99.

Reinke, W. M., Herman, K. C., & Newcomer, L. (2016). The Brief Student–Teacher Classroom Interaction Observation: Using dynamic indicators of behaviors in the classroom to predict outcomes and inform practice. *Assessment for Effective Intervention*, *42*, 32–42.

Riggs, C. G., & Mueller, P. H. (2001). Employment and utilization of paraeducators in inclusive settings. *Journal of Special Education*, *35*, 54–62.

Roberts, M. W., McMahon, R. J., Forehand, R., & Humphreys, L. (1978). The effect of parental instruction-giving on child compliance. *Behavior Therapy*, *9*, 793–798.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*, 493–529.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). Macmillan.

Sainsbury, M., & Sizmur, S. (1998). Level Descriptions in the National Curriculum: what kind of criterion referencing is this? *Oxford Review of Education*, *24*, 181–193.

Scull, J., & Winkler, A. M. (2011). Shifting trends in special education. Thomas B. Fordham Institute.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research*, *95*, 323–332.

Sobeck, E. (2016). *The effects of didactic instruction and performance feedback on paraeducators' use of positive behavior support strategies in inclusive settings* (Doctoral dissertation). University of Pittsburgh.

Spear, C. F. (2015). *Examining The relationship between implementation and student outcomes: The application of an implementation measurement framework*. University of Oregon. [Unpublished doctoral dissertation].

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*, 72–110.

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, *11*, 402–415.

Szadokierski, I., Burns, M. K., & McComas, J. J. (2017). Predicting intervention effectiveness from reading accuracy and rate measures through the instructional hierarchy: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *46*, 190–200.

U.S. Department of Education, National Center for Education Statistics, Statistics of State School Systems. (2017). Table 213.10: Staff employed in public elementary and secondary school systems, by type of assignment: Selected years, 1949–50 through fall 2015. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of education statistics* (2016 ed.). http://nces.ed.gov/programs/digest/current_tables.asp

U.S. Department of Labor, Bureau of Labor Statistics. (2016). Teacher assistants. *Occupational outlook handbook*. https://www.bls.gov/ooh/education-training-and-library/teacher-assistants.htm

Umbreit, J., Lane, K. L., & Dejud, C. (2004). Improving classroom behavior by modifying task difficulty: Effects of increasing the difficulty of too-easy tasks. *Journal of Positive Behavior Interventions*, *6*, 13–20. https://doi.org/10.1177/10983007040060010301

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*, 508–528.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education*, *27*, 365–378.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*, 357–364.

Wentzel, K. R. (2010). *Students' relationships with teachers. In Handbook of Research on Schools, Schooling and Human Development* (pp. 93–109. Routledge.

Werts, M. G., Zigmond, N., & Leeper, D. C. (2001). Paraprofessional proximity and academic engagement: Students with disabilities in primary aged classrooms. *Education and Training in Mental Retardation and Developmental Disabilities*, *36*, 424–440.

White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis*, *8*, 367–372.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)* (pp. 1–57). Ottawa: National Defense Headquarters.

How to cite this article: Lekwa AJ, Reddy LA. Current status and future directions in assessment of paraprofessional practices. *Psychology in the Schools*. 2021;1–21. https://doi.org/10.1002/pits.22480