

Running head: SWIFT-FIT PRELIMINARY TECHNICAL CHARACTERISTICS

**Development and Preliminary Technical Adequacy of the Schoolwide Integrated
Framework for Transformation Fidelity of Implementation Tool**

Bob Algozzine, Holly Morsbach Sweeney, Jeong Hoon Choi, Rob Horner, Wayne Sailor, Amy
B. McCart, Allyson Satter, Kathleen Lynne Lane

Author Note

Bob Algozzine, Department of Educational Leadership, University of North Carolina at Charlotte; Holly Morsbach Sweeney, Hoon Choi, Wayne Sailor, Amy B. McCart, Allyson Satter, Kathleen Lynne Lane, SWIFT Center, University of Kansas; Rob Horner, University of Oregon.

The authors produced this report under U.S. Department of Education, Office of Special Education Programs Grant No. H325Y120005, Schoolwide Integrated Framework for Transformation (SWIFT) Center. OSEP Project Officers Grace Zamora Durán and Tina Diamond served as the project officers. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service or enterprise mentioned in this publication is intended or should be inferred.

Correspondence concerning this article should be addressed to Bob Algozzine, EDLD, UNC Charlotte, Charlotte, NC 28223. E-mail: ralgozz@uncc.edu

Abstract

U.S. public education systems are required to provide free appropriate public education to students with disabilities in least restrictive environments that are appropriate to meet their individual needs. The practice of educating students with disabilities in neighborhood schools in age-appropriate general education classrooms and other school settings to meet this requirement has come to be known as “inclusive education.” The longstanding interest in keeping students with disabilities in the same classrooms with their neighbors and peers has created a need for reform to establish equity in America’s schools. Schoolwide Integrated Framework for Transformation (SWIFT) is a whole-system school reform model provided through a national technical assistance center that addresses core features of inclusive education support for elementary and middle schools, particularly those that are chronically low performing and those serving students with the most extensive needs. We describe the development and preliminary technical adequacy of Schoolwide Integrated Framework for Transformation Fidelity of Implementation Tool (SWIFT-FIT) as a means to document the extent to which schools are implementing inclusive education. Findings provide preliminary support for trained assessors using SWIFT-FIT as a valid and reliable instrument to produce evidence that describes the extent to which schools install, implement, and sustain these evidence-based practices. Researchers and other school personnel can use these data to evaluate the impact of implementation on progress as well as important student and other outcomes.

Keywords: inclusion, inclusive education, school reform, treatment integrity, assessment

Development and Preliminary Technical Adequacy of the Schoolwide Integrated Framework for
Transformation Fidelity of Implementation Tool

With the passage of the Education for All Handicapped Children Act (Public Law 94-142) in 1975, and with all its subsequent amendments and reauthorizations (e.g., Individuals with Disabilities Education Act or IDEA), states, local educational agencies or districts, and schools are required to provide a free appropriate public education to students with disabilities in the least restrictive environment that is appropriate to meet their individual needs. In principle, this requirement means that students with disabilities should have an opportunity to be educated with peers without disabilities; they should have access to the same curriculum or any other program as their peers without disabilities; and they should be provided supplementary aids and services necessary to achieve their individualized educational goals (McLeskey, Waldron, Spooner, & Algozzine, 2014; Sailor, 2014; Ysseldyke & Algozzine, 2006; Ysseldyke, Algozzine, & Thurlow, 2000).

In recent years, about 95% of these students were educated in general/regular education classrooms for at least some portion of the school day. However, 60% of these students were educated in a regular class 80% or more of the day; 20% were educated in a regular class 79% to 40% of the day; 14% were educated a regular class less than 40% of the day; and only about 5% were educated outside of a regular classroom in “other environments” (U.S. Department of Education, 2014). Although the term is not prominently used in federal documents on the implementation of IDEA, the practice of educating students with disabilities in the same classrooms and other environments as their peers has come to be known by many as “inclusive education” (Sailor, 2014). Despite “decades of advocacy” for education in general education classrooms, “there is tremendous variability of educational placement across the United States”

for students with disabilities and “most students with developmental disabilities continue to spend the majority of the school day in self-contained special education classrooms” (Brock & Schaefer, 2015, p. 54). Advocates of inclusive education are concerned with the equitable distribution of educational resources to all students; in this context, some believe that extensive reform must occur in the traditionally organized school to actualize inclusive education that is both academically and socially effective (Artiles & Kozleski, 2007; McCart, Sailor, Bezdek, & Satter, 2014; McLeskey et al., 2014).

Schoolwide Integrated Framework for Transformation and Fidelity Measure

Schoolwide Integrated Framework for Transformation (SWIFT) is a whole-system school reform model provided through a national technical assistance center (www.swiftschools.org) that addresses six critical issues facing America’s elementary and middle schools, particularly those that are chronically low performing. These issues include: (a) fragmented supports and lack of family engagement; (b) achievement gaps; (c) student engagement and behavior that impedes learning; (d) lack of implementation with fidelity of evidence-based interventions; (e) lack of sustainability and replication; and (f) lack of knowledge sharing and resource availability. SWIFT is multidimensional (see Figure 1) and its five domains of influence and core features are supported by evidence-based practice (McCart et al., 2014):

- **Administrative Leadership**
 - **Strong and Engaged Site Leadership:** Strong building-based leadership is a key component for developing and sustaining inclusive school practices (Ainscow & Sandhill, 2010; Waldron & McLeskey, 2010).
 - **Strong Educator Support System:** Principals play an important role in developing inclusive school-based programs (Hoppey & McLeskey, 2010).

- Multi-Tiered System of Support (MTSS)
 - Inclusive Academic Instruction: A multi-tiered framework should guide instruction by using effective general education strategies with all students and increasing the level of support for some students based on needs identified through screening and progress monitoring (Copeland & Cosbey, 2008; Sailor, 2009a, 2009b).
 - Inclusive Behavior Instruction: Implementing a multi-tiered behavior framework of support resulted in decreases in office discipline referrals, suspensions, and disruptive behaviors and increases in pro-social behavior (Bradshaw, Mitchell, & Leaf, 2010; Sailor, Wolf, Choi, & Roger, 2009; Sailor, Zuna, Choi, Thomas, McCart, & Roger, 2006).
- Integrated Educational Framework
 - Fully Integrated Organizational Structure: Organizational structures are most effective when they allow all students who need additional supports to benefit from resources that otherwise would only available to segregated populations of students (Sailor, 2009a).
 - Strong and Positive School Culture: “[R]esearch ... has shown repeatedly that the culture of schools is a strong influence on academic achievement” (Sailor, 2009a, p. 250).
- Family and Community Engagement:
 - Trusting Family Partnerships: Student achievement is likely to be higher when trusting partnerships exist than when they do not (Goddard, Tschannen-Moran, & Hoy, 2001; Hoy & Tarter, 1997; Sweetland & Hoy, 2000).
 - Trusting Community Partnerships: “Research indicates that when a collective group

of school, family, and community stakeholders work together, achievement gaps decrease” (Bryan & Henry, 2012, p. 408).

- **Inclusive Policy Structure and Practice:**
 - **Strong LEA/School Relationship:** A strong and supportive relationship between individual schools and their districts is critical for school reform and effective, sustainable practice (McLaughlin & Talbert, 2003)
 - **LEA Policy Framework:** A policy framework that is fully aligned with inclusive reform initiatives and that removes barriers to successful implementation must exist at the school, district, state, and federal levels (Kozleski & Smith, 2009).

SWIFT Fidelity of Implementation Tool (SWIFT-FIT; Morsbach Sweeney et al., 2013) measures the extent to which school personnel are using inclusive educational practices that align with SWIFT domains and features. The definition of “inclusion” reflected in SWIFT-FIT is whole-school, equity-based (Kozleski, & Thorius, 2014) rather than placement-based; applies to all students, not just those with disabilities; and is concerned with organizing all supportive instructional resources available to a school in a way that maximally engages all students in the teaching/learning process (McCart et al., 2014; McLeskey et al., 2014). The purpose of this article is to document the multi-step collaborative and iterative development process and preliminary technical adequacy of this tool. Evaluating technical adequacy involved psychometric analyses of a sample of completed measures as well as completion of series of studies designed to provide support for SWIFT-FIT validity, reliability, and usability.

Developing SWIFT-FIT

We followed accepted and widely recommended stages for development of assessment instruments in counseling, education, psychology, and other social science areas (Angoff, 1988;

Dimitrov, 2012; Messick, 1995a,b; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003; Soukakou, 2012). Specifically, we used a collaborative team-based approach to: (a) identify and define the purpose of SWIFT-FIT; (b) identify technical features of interest; and (c) identify core content to be included in a development draft.

Purpose of SWIFT-FIT. SWIFT Center's leadership and evaluation teams identified a need for a tool to document fidelity of implementation in a context of current practice and existing measures. They were guided by the following definition (Blase & Fixsen, 2013; Fixsen, Naoom, Blase, Friedman, & Wallace, 2005): Fidelity of implementation is the extent to which an intervention is delivered as conceived and planned. They envisioned a tool with both formative (i.e., support decision making about installing or improving) and summative (i.e., reflect growth and maturation data on inclusive education practices) value in assessing implementation fidelity. Thus, SWIFT-FIT is intended for use in documenting the extent to which technical assistance needed to implement a school-wide framework for inclusive education. It assesses whether support was provided (i.e., Were interventionists properly trained?) as well as the extent to which the framework was implemented as intended (e.g., Were interventionists properly practicing?). In this context, school personnel may use results of SWIFT-FIT assessments to document developmental practices, to evaluate the extent to which they are currently implementing inclusive education features, to plan training and technical assistance to improve their implementation, and to objectively document ongoing adoption of inclusive education over time (Morsbach Sweeney et al., 2013). Researchers may use the tool when evaluating the impact of inclusive education on measures of student achievement and related indicators of school success.

Technical features. Development and review teams led by persons with extensive experience in their respective fields and considered content experts assisted in identifying the technical features of SWIFT-FIT. As the first step, the development team explored extant and supportive knowledge about essential domains and features of effective inclusive education. Members of the leadership team then reviewed the SWIFT-FIT framework and provided feedback. Next, the development team constructed initial items by identifying detailed content for each feature to serve as a basis for documenting fidelity and collected feedback from potential users who were members of seven multi-organizational teams that comprise the SWIFT National Leadership Consortium (NLC).

Feedback was summarized and incorporated into an initial version of SWIFT-FIT with 53 items. These items were further reviewed by education professionals from inclusive schools (e.g., principals, teachers) for feedback on the practical importance and appropriateness of the scoring criteria. Information from all expert reviews contributed to iterative revisions until consensus was reached. Figure 2 illustrates two items from the tool.

Several versions of SWIFT-FIT were developed to refine its content and organization during a pilot and assessor training period that took place at schools in the Kansas City area. Between the second and third assessor trainings, data sources (i.e., interviews, document review, and observation) were refined to capture better evidences for scoring each item. This process of literature review, expert analysis, and consensus is considered to have established the best combined knowledge to serve as a basis for SWIFT-FIT technical features.

Core content. The final version of SWIFT-FIT includes 51 items across domains of Administrative Leadership ($n = 8$), Multi-Tiered System of Support ($n = 18$), Integrated Educational Framework ($n = 8$), Family and Community Engagement ($n = 7$), and Inclusive

Policy Structure and Practice ($n = 10$) and items ranging from 2 for Trusting Community Partnerships to 12 for Inclusive Academic Instruction and 3-6 for other features (see Figure 3). No weighting of items is included in any SWIFT-FIT score calculation. The tool assesses variation in the extent to which inclusive education is implemented (i.e., fidelity) using Likert-type item scores (i.e., ratings of 0, 1, 2, 3) of information provided by key informant (e.g., administrators, teachers, students, and parent) interviews, school and classroom observations, and reviews of artifacts and permanent products. Results are summarized in (a) a total score defined as the mean of the feature scores (average percentage of feature scores), (b) domain scores (percentage of points per domain), (c) individual feature scores (percentage of points per feature), and (d) an item summary that indicates the score per item and graphically displayed for review and continuing analysis. These scores provide a basis for evaluation and planning actions. The purpose of this article is to document preliminary evidence of the technical characteristics of SWIFT-FIT scores.

Method

We used data from several sources to document selected features of validity, reliability, and usability as preliminary evidence of SWIFT-FIT technical adequacy. We followed recommended and widely-used practices in counseling, education, psychology, and other social science areas (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, 2014; Angoff, 1988; Cronbach, 1988; Cronbach & Meehl, 1955; Dimitrov, 2012; Durlak & DuPre, 2008; Fixsen et al., 2005; Marion & Pellegrino, 2006; Messick, 1988, 1989, 1995a,b; Poggio, Glassnapp, Miller, Tollefson, & Burry, 1986; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003; Soukakou, 2012).

Settings and Participants

A total of 81 SWIFT-FIT assessments were completed from 81 schools across 11 states. We collected data from three settings: Knowledge Development Sites (KDS), assessor training sites, and SWIFT partner sites. As a part of a larger study, six KDS schools were selected as exemplary inclusive education sites evidencing positive outcomes for all students, including those with disabilities (Mitchiner, 2014; Shogren et al., 2015). First, using a key informant strategy, members of the NLC—which is comprised of a national group of researchers and technical assistance providers—nominated 37 schools based on their knowledge of each of the school’s demonstrated strengths in areas such as: inclusive educational practices, use of a multi-tiered system of support, strong family and community partnerships, and indicators of achievement growth of students (including those with disabilities). An initial screening of these areas, willingness to participate, and additional grant requirements reduced the sample to 30 schools. Next, surveys and phone interviews were conducted to gather information about school composition and more in-depth understanding of ways in which they implemented key features of inclusive education. The leadership team then reviewed this information and weighed considerations related to geographic location, population demographics and ages served, to narrow the sample to 11 schools. Finally, five researchers and technical assistance providers visited schools for one-day, conducting guided interviews and assessing inclusive practices. Data were summarized; strengths and weaknesses considered; and, ultimately, six schools were selected as KDS where further study would take place. Each KDS had evidence of some, but not all, core features reflected in SWIFT-FIT (Shogren et al., 2015).

The second setting was Kansas City area schools, where trainers and assessors in training learned to conduct SWIFT-FIT assessments. We made request of 22 schools based on our previous relationships with district staff, principal, or school staff. Nine schools initially agreed

to participate; however, one school opted out due to scheduling conflicts. The remaining eight schools participated as data sources in SWIFT-FIT evaluation process, and are referred to here as Cohort sites.

The third setting was SWIFT partner schools. Personnel from each of SWIFT's participating State Educational Agencies (SEA; Maryland, Oregon, Mississippi, Vermont and New Hampshire) selected two districts or local educational agencies (LEAs) to identify individual schools to serve as SWIFT partner sites. SWIFT partners participate in technical assistance for installing and sustaining the SWIFT educational framework (Figure 1). At the time of this psychometric evaluation study, Maryland, Oregon and Mississippi SEAs each chose 16 schools; Vermont SEA chose 11 schools and; New Hampshire SEA chose 8 schools, for a total of 67 schools comprised of elementary, middle and/or combined K-8th schools located in varying areas across each state, with no more than five sites in any one district. Partner sites are in urban (21%), suburban (6%) and rural (73%) communities and vary in size (average student enrollment = 397, $SD = 182$, Range = 53–938) and demographics.

Procedure

We focused on content, construct, and concurrent validity. We also documented the extent to which scores were similar for subsets of items on a single administration (i.e., internal consistency), the extent to which scores were similar across repeated administrations (i.e., rater agreement), and the extent to which users provided favorable ratings of SWIFT-FIT.

SWIFT-FIT assessors were recruited from the SWIFT Center, academic community, and educational agencies in states and districts that were implementing SWIFT. Qualification and experience requirements of the assessors were clearly identified by the development team. In order to meet the qualification criteria, assessors were required to have an educational

background, including special education and related areas (e.g., psychology, communication disorders, and child development), and/or work experience in schools or other related agencies. One-day intensive training was provided to selected assessors. The training included instruction on SWIFT domains and features, review of each item, interpretation of rubric criteria, review of the assessment procedure and related materials such as interview questions, required document review, and scoring practice with mocked scenario (i.e., transcribed interviews and document examples).

SWIFT-FIT requires approximately 5 to 6 hours to administer and 1 to 2 hours to score. Assessors review documents as evidence of the items within the SWIFT framework. They conduct individual and/or focus group interviews with school stakeholders, which include the leadership team members, principal, staff, educators, and family and community partners. Assessors also conduct observations in classrooms and common areas of the school. These data sources (i.e., interviews, observation, and document review results) are used for assessors to score all items across the five domains and 10 core features of SWIFT-FIT, and a detailed scoring rubric is provided for each item (see Figure 3). Domain and feature scores are generated as a percentage of total possible points for the corresponding domain and feature. Total score is an average percentage of feature scores.

Analysis

To evaluate the technical adequacy of SWIFT-FIT, we conducted (a) calculation of mean and standard deviation and content validity index (CVI) analyses for examining content validity, (b) effect size analysis for examining construct validity, (c) correlational analyses for examining concurrent validity, (d) analyses of observer agreement for examining inter-rater reliability, (e) correlational analyses and Cronbach's alpha tests for examining internal consistency of

reliability, and (f) analysis of survey results for examining usability of SWIFT-FIT. Detailed methods and results of the analyses are presented below in the results section.

Results

Validity

For content-related evidence, we documented the extent to which scores on SWIFT-FIT “looked like” acceptable and reasonable indicators of implementation (i.e., face validity) and the extent to which subject matter experts (SMEs) rated items and the overall measure as adequate and appropriate indicators of levels of implementation. For construct-related evidence, we documented the extent to which scores on SWIFT-FIT reflected naturally-occurring variation in implementation of inclusive practices. For concurrent validity, we documented the extent to which scores on SWIFT-FIT were related to scores on a similar measure administered at the same time.

Content validity. We documented the extent to which SWIFT-FIT “looked like” an acceptable and reasonable indicator of inclusive education by comparing its content and that of a similar measure used to assess implementation levels of the Schoolwide Applications Model (SAM: Sailor & Roger, 2005), from which the SWIFT framework for inclusive education emerged. Schoolwide Applications Model Analysis of Selected Critical Features (SAMAN; Sailor & Roger, 2003, 2008) was created as the fidelity tool to monitor implementation fidelity of SAM with assessment items for 15 critical features using a 4-point rubric for each. Lessons learned from work with the SAM model of inclusive education plus new knowledge gained from the KDS were later included into the SWIFT framework, and subsequently into SWIFT-FIT for measurement. SWIFT domains and features, although predicated on the earlier SAM work, are

sufficiently different from the SAMAN tool features to require development of the present fidelity of implementation tool.

Much was learned from earlier SAM implementation experiences that was incorporated into the SWIFT framework, such as the importance of (a) competent classroom instruction and ensuring a strong educator support system, (b) valuing a positive school climate and culturally appropriate and responsive practices, (c) building relationships between the LEA and school, and (d) acknowledging importance of LEA policy structure to support sustainable change. These features are unique to SWIFT-FIT. However, more than 80% agreement was evident for content found in both SWIFT-FIT and SAMAN related to strong and engaged site leadership, inclusive academic and behavior instruction, fully integrated organizational structure, and family and community engagement.

Overall, SWIFT-FIT “looks like” an acceptable and reasonable indicator of an inclusive educational framework. Additional analyses supported its content-related validity. Using an online survey, we documented the extent to which SMEs rated SWIFT-FIT content as adequate and appropriate indicators of levels of implementation. We documented the consistency of ratings using both narrative and statistical summaries.

Twenty-five representatives from KDS were invited to participate as experts: 6 (24%) were administrators, 18 (72%) were instructional staff members, and 1 (4%) was a parent. Although national experts would also provide important insights, we reasoned that the opinions of people who were “living” the indicators in their daily work and achieving successful outcomes with inclusion would provide a valuable contribution to the validity of SWIFT-FIT.

The survey prompted participants to rate each item on SWIFT-FIT, reflecting their opinion of its importance using a 4-point Likert-type scale ranging from 1 = *not very important*

to 4 = *very important*. Additionally, using an open-response option, they were asked to provide examples of school-based indicators that could be used as evidence for each of the 10 features. We sent a survey link to each participant, explaining the reason they were chosen for participation, stating the purpose of the survey, and requesting their participation. A week later, another email thanked those who had participated and requested those who had not yet responded to please take the survey. Results from the online software program, Qualtrics, indicated that the survey took an average of 12-15 minutes to complete and had a 76% response rate (19 of 25 participants).

Expert ratings of importance across SWIFT domains and features are summarized in Table 1. The average overall rating was 3.66 ($SD = 0.29$); and all domains and features had average ratings indicative of *important* to *very important*. Domain means ranged between a low ($M = 3.48$) for Inclusive Policy Structure and Practice to a high ($M = 3.85$) for Integrated Educational Framework. Average ratings for features ranged between a low ($M = 3.21$) for Trusting Community Partnerships to a high ($M = 3.86$) for Strong and Positive School Culture. Note, however, that the *low* averages still reflect *high* ratings of importance on the 4-point, Likert-type scale.

Six (12%) of the 51 items were rated as not very important or somewhat important by 10% or more of participants: instructional coaching (19%), progress monitoring for math (10.5%), universal screening for behavior (16%), and all three items related to community partnerships at the school and LEA level (26%, 16% and 11%, respectively). The lower rating for universal screening for behavior is likely due to a commonly reported lack of experience with and knowledge about the practice (Lane et al., 2009; Lane, Menzies, Bruhn, & Crnabori, 2011; Lane, Menzies, Kalberg, & Oakes, 2012; Lane, Menzies, Oakes, & Kalberg, 2012). For

example, KDS interviews reflected unfamiliarity with this practice and few used systematic tiered supports for behavior, because they believed that high levels of academic engagement and added supports in each classroom provided proactive climates that resulted in “few behavior challenges” and reduced the need for a focus on them.

The highest rated items were those within the Integrated Educational Framework domain, which is the core of inclusive educational practices. Two items rated as very important by all respondents were working collaboratively to monitor and plan academic interventions, and to collaboratively teach all students in fully inclusive classrooms.

To obtain a statistical summary of content validity, we randomly selected 10 ratings for further analysis. Representing the proportion of experts who rate a measure’s items as relevant, important, or reflective of the content being measured, the content validity index (CVI) is a widely reported measure of the content-related validity in health and social sciences fields (Beck & Gable, 2001; Dimitrov, 2012; Polit & Beck, 2006; Rubio et al., 2003; Lawshe, 1975). Lynn (1986) and Polit and Beck (2006) recommended that CVI should be 1.00 when ratings of less than six experts are compared, and no lower than 0.78 when six or more experts rate the items. We calculated CVI for each item by counting the number of experts who rated it *very important* or *important* and dividing that number by 10, the total number of experts that we selected.

The average proportion of items that received ratings of *important* or *very important* on SWIFT-FIT was 0.94. The average CVI for domains of influence ranged from 0.93 (Administrative Leadership) to 0.98 (Integrated Educational Framework). The average CVI for features ranged from 0.87 (Strong Educator Support System) to 1.0 (Fully Integrated Organizational Structure and Trusting Family Partnerships).

Evidence from our content-related analyses documented the extent to which SWIFT-FIT “looked like” an acceptable and reasonable indicator of implementation and the extent to which expert panel members rated items and the whole measure as adequate and appropriate indicators of levels of implementation. Overall, evidence supports the content-related validity of SWIFT-FIT.

Construct validity. To document the extent to which scores on SWIFT-FIT reflected naturally-occurring similarities and differences in implementation, we compared scores for KDS and Cohort schools. We administered a pilot version of SWIFT-FIT in the six KDS and administered the final version in three of these schools approximately six months later. We also collected SWIFT-FIT scores in eight Cohort sites. We reasoned that scores on different occasions would be similar for KDS and that scores from schools engaged in different levels of inclusive practices would be different.

Means, standard deviations, and comparison statistics across different versions of SWIFT-FIT are in Table 2. Domain, feature, and total scores were statistically similar. SWIFT-FIT total scores for the pilot and final versions were not statistically significantly different. Effect sizes were small for Administrative Leadership, Multi-tiered System of Support, and Inclusive Policy Structure and Practice; and moderate for Integrated Educational Framework and Family and Community Engagement score differences. Similarly, small practical differences were evident for all features except Inclusive Behavior Instruction, Fully Integrated Organizational Structure, and Trusting Community Partnerships. In general, the observed similarities were greater than differences across the two versions of SWIFT-FIT.

Means, standard deviations, and comparison statistics across Cohort and KDS schools are in Table 3. SWIFT-FIT total scores for Cohort and KDS schools were statistically significantly

different. Statistically significant differences and large effect sizes were also evident for Administrative Leadership, Multi-tiered System of Support, and Integrated Educational Framework domain scores; however, Family and Community and Inclusive Policy Structure and Practice domain scores were statistically similar for Cohort schools and KDS, respectively. Large practical differences were evident for four features: Strong and Engaged Site Leadership, Inclusive Academic Instruction, Fully Integrated Organizational Structure, and Strong and Positive School Climate. Non-statistically significant, but large practical differences were evident for Strong Educator Support System, Trusting Family Partnerships, and LEA Policy Framework; and non-statistically significant and small practical differences were evident for Inclusive Behavior Instruction, Trusting Community Partnerships, and Strong LEA/School Relationship. Observed scores were higher for KDS than Cohort sites.

We expected that scores in some areas for the KDS, which are recognized as models of inclusive educational practices, would be high (Table 2), and they were, suggesting that domain and features “reflect” or “represent” reasonable implementation indicators. We also expected to document small effect sizes between the two SWIFT-FIT administrations, which we did for three of the five domains and seven of the 10 features. However, we noted moderate to high effect sizes for the three features for which the most changes to content between the first and second administration were made. Major changes included refinement of scoring rubric to more accurately reflect gradations of rating category. Some items were separated as independent items to deepen content and some items were combined to eliminate redundancy. Inclusive Behavior Instruction ($d = 0.66$) content and criterion for scoring became more expansive and detailed in the second version. Correlated with this change, the KDS scored low on the second administration; an increase in expectations would result in lower overall mean scores. Another

possible reason for lower scores for this feature is that, anecdotally, some KDS commented that they did not believe that they needed formalized behavior structures in place due to having high levels of student academic engagement and low rates of problem behavior.

The content for Fully Integrated Organizational Structure ($d = 0.89$) had significant changes between administrations with six of its eight items revised. Lessons learned in this area from KDS personnel during the first administration (which also served as pilot testing in the development process of SWIFT-FIT) resulted in improvements to the criterion for these items, which would show as improvements in their results. These items would more closely represent the suggestions for improvement that matched their existing practices.

The Trusting Community Partnership feature ($d = 1.33$) was also revised between administrations. Input from national experts into this feature were considered and revision made between administrations, resulting in a construct of community partnerships not commonly seen in schools to date, yet set as expectations for the future. As a result, the mean score for this feature decreased between administrations.

Concurrent validity. To document the extent to which scores are related to those for similar or different measures administered at the same time, two trained and reliable assessors for SWIFT-FIT and SAMAN used both tools to rate the same school. The process of administering these measures was similar enough to use the evidence gathered from schools for SWIFT-FIT (i.e., classroom observation, interviews with stakeholders, and document review) to concurrently score SAMAN. However, SAMAN results would have been insufficient to score SWIFT-FIT's additional depth and breadth of evidence. The results from our face validity study comparison examining the similarity of items between SWIFT-FIT and SAMAN support this

decision. Seven schools (three KDS and four SWIFT partner schools) were dually scored using SWIFT-FIT and SAMAN.

Concurrent validity of SWIFT-FIT was evident in positive associations between the total score on the SAMAN and SWIFT-FIT domain, feature, and total scores. Statistically significant correlations were evident for SWIFT-FIT total score (.70), and Administrative Leadership (.71), Multi-Tiered System of Support (.74), Integrated Educational Framework (.88), and Family and Community Engagement (.63) domain scores. A low negative association (-.18) was documented for Inclusive Policy Structure and Practice domain and total score on SAMAN. Statistically significant correlations were also evident for five features: Strong and Engaged Site Leadership (.77), Inclusive Academic Instruction (.82), Fully Integrated Organizational Structure (.93), Strong and Positive School Culture (.72), and Trusting Family Partnerships (.69). Moderate non-statistically significant relations were observed for two features: Strong Educator Support System (.55) and Trusting Community Partnerships (.42). Low relationships were evident for two features: Inclusive Behavior Instruction (.19) and LEA Policy Framework (.03). A moderate negative relationship was evident for Strong LEA/School Relationship (-.33).

The strongest associations would be expected in areas that were similar between the SAM and SWIFT frameworks, for which these tools are used: focused leadership, academic and behavior tiered supports, integrated educational structures, integrated educational practices (found in the Strong and Positive School Culture feature of SWIFT), and family and community involvement. These expectations were validated for all but two areas: behavior tiered supports and community involvement. The moderate non-statistically significant results for the two features Strong Educator Support System (.55) and Trusting Community Partnerships (.42) and low relationship for Inclusive Behavior Instruction (.19) are likely due to the specific details that

were built into the SWIFT framework from lessons learned during SAM implementation and the KDS study. The low or negative correlations found in Inclusive Policy Structure and Practice domain ($-.18$) and its two features, LEA Policy Framework ($.03$) and Strong LEA/School Relationship ($-.33$), are to be expected. When the SWIFT framework was built, it added a focus on LEA level policies and additional expectations surrounding the working relationship with the school, which were not present in SAM.

Reliability

For test-retest evidence, we documented the extent to which scores were similar across repeated administrations of SWIFT-FIT. For estimates of internal consistency, we documented the extent to which scores were similar for subsets of items on a single administration of SWIFT-FIT. For evidence of rater agreement, we compared SWIFT-FIT scores for different assessors.

Internal consistency. To document internal consistency reliability estimates we calculated Cronbach's alpha (α) (Cronbach, 1951; Cronbach, 1969; Cronbach & Meehl, 1955) across partner schools for five domains, 10 features, and 51 items of SWIFT-FIT. Cronbach's alpha provides a measure of the internal consistency with a number between 0 and 1, and the acceptable values vary in different reports, ranging from $.70$ to $.95$. Correlations between each item with its corresponding feature score (r_{fs}) and between each item and the total SWIFT-FIT score (r_{tot}) were also calculated (Cohen, Kincaid, & Childs, 2007).

In general, reliability of SWIFT-FIT scores was acceptable and above the general cutoff of interest ($.70$) for development purposes (Henson, 2001). Cronbach's alphas for total SWIFT-FIT, Administrative Leadership, Multi-Tiered System of Support, Integrated Educational Framework, Family and Community Engagement, and Inclusive Policy Structure and Practice domains were $.96$, $.88$, $.92$, $.69$, $.84$, and $.84$, respectively. Cronbach's alphas for features ranged

from .44 to .92. Fully Integrated Organizational Structure (4 items) and Trusting Community Partnerships (2 items) had the lowest internal consistency ($\alpha = .44$ and $.47$, respectively), and Strong and Positive School Culture (4 items) was slightly below ($\alpha = .67$) the target value of $.70$.

Item-total correlations revealed moderate to strong correlation between items and corresponding feature total scores. The overall r_{ss} score range was from $.37$ to $.91$. The correlation between each item and overall total score (r_{tot}) showed above moderate correlations for most items, while some items revealed moderate or weak relationships.

Rater agreement. We used data from 11 assessors and 13 schools to document scoring consistency. Participants included primary assessors, who had met criterion for becoming a trained and reliable assessor, and trainees who were not yet at criterion. Both practice schools and partner schools were included. Practice schools agreed to participate in training assessors and are not part of SWIFT Center technical assistance partnerships. Partner schools were those working with SWIFT Center to implement inclusive education.

For purposes of this preliminary study, scores between trained and reliable assessors and assessors in training were calculated and reported. In each administration one data collector served as the primary assessor while a second, and in some cases, a third person observed the process and also scored SWIFT-FIT. Observers were assessors in training and had received preliminary training on SWIFT domains and features as well as SWIFT-FIT administration protocols. Some observers were conducting their first session with a trained assessor, and others their second or third observation. Inter-observer agreement was based on an item-by-item comparison and calculated by dividing the number of items with perfect agreement by the total number of SWIFT-FIT items and multiplying by 100. On the occasions when three assessors conducted a SWIFT-FIT assessment for one school (one primary assessor and two observers),

inter-observer agreement was calculated between each pair, resulting in 30 pairings for analysis. We report the inter-observer agreement between trained assessors and assessors in training to provide preliminary evidence as to the reliability of the tool.

The average inter-observer agreement on SWIFT-FIT items across the 13 schools and 30 possible pairings of assessors was 79.6% (Range = 60.0%–96.1%). Excluding the inter-observer agreement calculations between two observers (on those occasions where there was a primary assessor and two observers) the average was 82.3% (Range = 62.0%–96.1%).

Individual item inter-observer agreements ranged from 53.3% to 100%. Agreements from 19 items were below 80%, and 10 items had less than 70% of inter-observer agreements. Average inter-observer agreements of three features, Strong LEA/School Relationship ($M = 68.7\%$, $SD = 8.4\%$, Range = 56.7%–80%), LEA Policy Framework ($M = 73.3\%$, $SD = 17.0\%$, Range = 53.3%–93.3%), and Trusting Family Partnerships ($M = 75.3\%$, $SD = 9.6\%$, Range = 63.3%–83.3%), were lower than 80%. Those items and features were intensively reviewed and their scoring criteria were refined for future training.

Usability

To document the extent to which assessors favorably rated SWIFT-FIT ease of use, an online survey was administered using *Qualtrics*. Participants rated each of 12 statements using a 4-point Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. Each trained assessor administered SWIFT-FIT a minimum of twice. Survey statements covered aspects of administering SWIFT-FIT as well as various sections of the tool. For example, the survey presented for rating such statements as: The steps in a SWIFT-FIT administration are described clearly in the protocol; Item descriptions were clearly written and easy to understand; and The process for completing SWIFT-FIT is reasonable, appropriate, and effective. Additionally, using

an open answer option, respondents commented on what they liked best about SWIFT-FIT; liked least about SWIFT-FIT; and what, if any, improvements would they suggest for SWIFT-FIT. A survey link was emailed to each participant to explain the reason they were chosen for participation, the purpose of the survey, and ask for their participation. A week later, another email thanked those who had participated and requested those who as yet had not to please take the survey. Results indicated that the survey took on average 12-15 minutes to complete and an 82% response rate (9 of 11) from participants.

The overall mean was 3.10 for items scored on the survey asking participants their level of agreement with the usability of SWIFT-FIT. All items had a mean indicative of *agree* or *strongly agree* ratings. Five items focused on clarity of specific aspects within SWIFT-FIT (purpose, steps, process for preparing schools, process for scoring and item descriptions). Respondents rated overall agreement of usability with these items, resulting in a mean of 3.17. Two survey items focused on how helpful supplemental information was for assessors in conducting SWIFT-FIT assessment: assessor notes, which are to be used to clarify intent of items on SWIFT-FIT, and a list of positions occupied by persons to interview and associated questions. Respondents rated *agree* to *strongly agree* of usability with these items, resulting in a mean of 3.44. SWIFT-FIT assessors rated lower levels of agreement with the scoring rubrics being clearly written and easy to use ($M = 2.89$); the process for completing SWIFT-FIT as reasonable, appropriate, and effective ($M = 2.89$); and SWIFT-FIT as an effective or efficient assessment of implementation fidelity ($M = 2.67$ and 2.89 , respectively). Overall usability of SWIFT-FIT had a mean of 2.56. Our analysis indicated that assessors favorably rated the usability of SWIFT-FIT. Additionally, they provided valuable insight into possible improvements to the tool as well as the training protocol for SWIFT-FIT assessors. A revised

SWIFT-FIT training process was developed in response to feedback gathered from SWIFT-FIT assessors. The current training process includes three parts: (a) a series of online modules that provide an overview and examples for each feature, (b) a mock SWIFT-FIT that assessors score, and (c) on-site experience where assessors conduct a SWIFT-FIT alongside experienced assessors.

Discussion

Research is emerging on the school reform and systemic changes needed to ensure that students with disabilities are included in the same educational experiences as their natural neighbors and peers (McCart et al., 2014). It is generally agreed that definitions of inclusive practices put forth the following components (McLeskey et al., 2014, pp. 4-5):

- School provide comprehensive and ongoing support to better meet the need of a diverse range of students.
- Professionals work collaboratively to provide support and effective instruction.
- Students are educated in natural settings that are highly effective in meeting their needs.
- Students are educated together.
- Students are valued members of all classrooms.
- Students are provided supports to meet individual needs and achieve valued appropriate learner outcomes.

We used a collaborative team-based approach to develop, field test, and document the preliminary technical characteristics of a measure of the extent to which a school is implementing domains and core features of inclusive education. Our work complements what is known and provides evidence of the value of SWIFT-FIT for documenting inclusive practices.

The content of SWIFT-FIT adequately reflects the “big ideas” of evidence-based and effective inclusive education (Ainscow & Sandhill, 2010; Artiles & Kozleski, 2007; McCart et al., 2014; McLeskey et al., 2014; Sailor, 2009a, 2009b; Sailor, Wolf, Choi, & Roger, 2009; Sailor, Zuna, Choi, Thomas, McCart, & Roger, 2006; Waldron & McLeskey, 2010). This extant professional knowledge and body of evidence-based research is reflected in SWIFT-FIT assessment content and scores.

In content-related validity, SWIFT-FIT appeared to be a reasonable measure for assessing an inclusive educational framework when compared to other assessment tools used in a similar schoolwide model (i.e., face validity). Content similarity of SWIFT-FIT with SAMAN was 73.6%. Experts also rated SWIFT-FIT content as adequate indicators. The content validity index (CVI) for the expert validity revealed that the proportion of experts who rated SWIFT-FIT items as relevant and important was high, showing that the average for domains of influence ranged from 0.93 to 0.98, and the average for features ranged from 0.87 to 1.0. SWIFT-FIT total scores collected from KDS showed statistically significant differences when compared to Cohort sites, which is evidence that SWIFT-FIT reflects naturally-occurring similarities and differences in implementation (i.e., logical validity). All these findings support the technical adequacy of the tool to document variation in implementation of inclusive practices.

Second, SWIFT-FIT produces scores comparable to ones on similar measures of implementation of inclusive practices. The concurrent validity, which was examined through correlations between SWIFT-FIT and SAMAN, showed statistically significant and positive correlations across total scores (correlation coefficient, .70) and all domain scores (correlation coefficient ranged from .63 to .88) except Inclusive Policy Structure and Practice.

Third, SWIFT-FIT is a reliable tool producing internally consistent scores on similar items and similar results with different assessors. Internal consistency analysis results revealed high Cronbach's alphas for total SWIFT-FIT and all domain scores except Integrated Educational Framework, which had slightly lower Cronbach's alpha (.69) than the general cutoff for research (.70). Inter-rater agreement analysis results showed average agreement between two assessors was 82.3%.

Finally, trained and experienced assessors found SWIFT-FIT to be easy to understand, and administration process was reasonable, appropriate, and effective to complete. The usability survey average rating was 3.17 out of 4, which can be interpreted as favorable ratings of the usability.

Although other measures of inclusion-related practices exist (e.g., *Autism Program Environment Rating Scale-Preschool/Elementary* (APERS-PE): National Professional Development Center on ASD, 2011; *Inclusive Classroom Profile* (ICP): Soukakou, 2012; *School-Age Care Environment Rating Scale* (SACERS): Harms, Jacobs, & White, 1996), most represent selected assessments for targeted age- or disability-groups (e.g., preschool, autism) or settings (e.g., classrooms). Our analysis supports the value of SWIFT-FIT in broad areas of intervention implementation. SWIFT-FIT is used to document the extent to which five domains and 10 features of evidence-based practices that support inclusive education are evident in schools. The total, domain, and core feature scores of SWIFT-FIT are reviewed by state, district, and school teams along with outcome data such as student academic achievement, behavior performance, and school climate survey results. Collectively, the information is used to help teams identify and prioritize opportunities for growth over the next 6 to 12 months of SWIFT implementation. The intent of this priority and practice planning process is to help a school or

district cull and incorporate existing school improvement plans into the core features of the SWIFT framework. The process also helps schools identify which action steps can have the biggest impact with the smallest effort. Our findings provide preliminary support for using SWIFT-FIT as a valid and reliable formative and summative data source for use in planning.

Limitations

Several limitations of the study are important to consider. Although we found adequate association between SWIFT-FIT and SAMAN, further studies to examine the concurrent validity between SWIFT-FIT and other assessments measuring effective inclusive practices are merited. SAMAN was compared with SWIFT-FIT in the current study because it was the fidelity of implementation tool for SAM, the predecessor to SWIFT for implementing inclusive education. In this regard, however, it is important to note that measures of inclusive practices are not widely available and, with continuing use and refinement, SWIFT-FIT may become the gold standard for addressing this important focus of educational reform.

Inter-rater reliability needs to be examined with more samples and a broader range of practices. Although scores from paired trained assessors were evaluated in the current study, they were collected before our training was clearly defined and a manual fully developed. Current SWIFT-FIT training procedure for new assessors includes online content training and field training. In future studies, scores reflecting inter-rater agreement between fully trained assessors need to be compared.

It is also important to acknowledge that SWIFT-FIT is a measure of broad educational practices. Some areas of focus are implemented within school and classroom contexts (e.g., Strong and Engaged Site Leadership, Strong Educator Support System). Others are more defined and controlled by external agents and agencies (e.g., Strong LEA/School Relationship

and LEA Policy Framework); and inter-rater agreement for these areas of practice were lower than those for school-based features. Scores on SWIFT-FIT are based on information gathered from review of extant documents as well as from individual and team interviews and targeted classroom observations. While comprehensive in intent, the scope of the measure and the data used to inform scoring create challenges for consistently collecting and documenting information across raters. Our review of the range and limits of obtained inter-rater reliability was concerning; however, it also was informative and directive of improvements subsequently included in training protocols and continuing data collection.

SWIFT is currently implemented in 64 schools in five states, and data for the current reliability study came from these schools. A sample size of 64 is relatively small to generalize our findings to other schools and other states. Sample diversity also needs to be considered in future studies. All 64 schools in the current study were purposively selected by five states to represent low-performing and high-need schools only. The large number and variability in school samples will positively impact reliability and validity in the future research. Additional longitudinal research needs to be planned and conducted to examine the extent to which scores on SWIFT-FIT are related to enhanced achievement on the part of all students and all subgroups, as well as increased inclusive education.

Finally, although SWIFT-FIT is administered by trained, reliable assessors, we recognized schools can “game the system” by “putting their best foot forward” on assessment day. This limitation is, to some degree, inevitable for all such external assessments, but SWIFT-FIT reliance on classroom observations coupled with document review by assessors is felt to adequately address shortcomings of information obtained from interviewees. A multiple data source approach provides a reasonably objective overall picture school progress implementing

inclusive educational practices. We expect the revised SWIFT-FIT assessor training would affect scoring accuracy and efficiency of administration through scenario-based examples and on-site training with experienced assessors. Administration requirements including school preparation and communication were included in the revised SWIFT-FIT training. In order to minimize time constraints and increase usability in practice settings, further research should consider the relationships among the revised training and effective administration (e.g., time requirements, collection of evidence, review process). The use of SWIFT-FIT data in the action planning process and their relationships to student outcomes also need to be studied to weigh the effects of each SWIFT domain and feature, and each SWIFT-FIT item.

Despite these limitations, SWIFT-FIT data play a critical role in guiding decision-making throughout SWIFT technical assistance and continuous improvement processes. SWIFT-FIT data are used along with other data sources to identify priorities and determine action steps needed to generate improvement of student outcomes for partner schools and districts. Score improvements on features are reviewed annually to investigate implementation progress and fidelity.

Conclusion

SWIFT-FIT was designed to be an efficient measure of implementation fidelity by documenting inclusive education status at the school level. Use of the tool provides an effective means of reliably and validly assessing core aspects of inclusive education. It yields domain and feature scores as well as total scores. The findings from SWIFT-FIT can be reviewed and discussed with a school leadership team and used to plan and prioritize implementation actions. Aggregate data can also be used by districts and state educational agencies to monitor implementation of inclusive education practices, and more importantly, provide appropriate

supports for schools. As such, SWIFT-FIT provides a useful tool for educators installing evidence-based inclusive education practices for all students and provides researchers with a reliable and valid tool with which to estimate the impact of installation of features of the SWIFT framework on measured student academic success. Although further research needs to be done to expand on these results, the current study provides preliminary evidence that SWIFT-FIT adequately assesses the level of inclusive education practices and SWIFT implementation fidelity.

References

- Ainscow, M., & Sandhill, A. (2010). Developing inclusive education systems: the role of organisational cultures and leadership. *International Journal of Inclusive Education, 14*, 401-416. doi:10.1080/13603110802504903
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. I. Braun (Eds), *Test validity* (pp. 19-32). Hillsdale, NJ. Lawrence Erlbaum Associates, Inc.
- Artiles, A., & Kozleski, E. (2007). *Beyond convictions: Interrogating culture, history, and power in inclusive education*. *Language Arts, 84*, 351-358. doi:10.1177/0042085909337595
- Beck, C. T., & Gable, R. K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement, 9*(2), 201-215.
- Blase, K., & Fixsen, D. (2013, February). *Core intervention components: Identifying and operationalizing what makes programs work*. ASPE Research Brief. Retrieved from <http://nirn.fpg.unc.edu/resources/core-intervention-components>
- Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of school wide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12*, 133-148. doi:10.1542/peds.2012-0243

- Brock, M. E., & Schaefer, J. M. (2015). Location matters: Geographic location and educational placement of students with developmental disabilities. *Research & Practice for Persons with Severe Disabilities, 40*, 154-164. doi:10.1177/1540796915591988
- Bryan, J., & Henry, L. (2012). A model for building school-family-community partnerships: Principles and process. *Journal of Counseling and Development, 90*, 408-420. doi:10.1002/j.1556-6676.2012.00052.x
- Cohen, R., Kincaid, D., & Childs, K. (2007). Measuring school-wide positive behavior support implementation: Development and validation of the Benchmarks of Quality (BoQ). *Journal of Positive Behavior Interventions, 9*, 203-213. doi:10.1177/10983007070090040301
- Copeland, S. R., & Cosbey, J. (2008). Making progress in the general curriculum: Rethinking effective instructional practices. *Research and Practice for Persons with Severe Disabilities, 33*, 214-227. doi:10.2511/rpsd.33.4.214
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi:10.1007/BF02310555
- Cronbach, L. J. (1969). Validation of educational measures. *Proceedings of the 1969 Invitational Conference on Testing Problems* (pp. 35-52). Princeton, NJ: Educational Testing Service.
- Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H. I. Braun (Eds), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302. doi:10.1037/h0040957

- Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.
doi:10.1007/s10464-008-9165-0
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). Retrieved from <http://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature>
- Goddard, R. D., Tschannen-Moran, M., & Hoy, W. K. (2001). A multilevel examination of the distribution and effects of teacher trust in students and parents in urban elementary schools. *The Elementary School Journal, 102*(1), 3-17.
- Harms, T., Jacobs, E., & White, D. (1996). *School Age Care Environment Rating Scale*. New York, NY: Teachers College Press.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177-189.
- Hoppey, D., & McLeskey, J. (2010). A case study of principal leadership in an effective inclusive school. *The Journal of Special Education, 46*, 245-256.
doi:10.1177/0022466910390507

- Hoy, W. K., & Tarter, C. J. (1997). *The road to open and healthy schools: A handbook for change*. Thousand Oaks, CA: Corwin Press.
- Kozleski, E. B., & Smith, A. (2009). The complexities of systems change in creating equity for students with disabilities in urban schools. *Urban Education, 44*, 427-451.
doi:10.1177/0042085909337595
- Kozleski, E. B. & Thorius, K. K. (Eds.). (2014). *Ability, equity, and culture: Sustaining inclusive urban education reform*. New York: Teachers College Press.
- Lane, K. L., Little, M. A., Casey, A. M., Lambert, W., Wehby, J. H., Weisenbach, J. L., & Phillips, A. (2009). A comparison of systematic screening tools for emotional and behavioral disorders: How do they compare? *Journal of Emotional and Behavioral Disorders, 17*, 93-105. doi:10.1177/1063426609341069
- Lane, K. L., Menzies, H., Bruhn, A., & Crnabori, M. (2011). *Managing challenging behaviors in schools: Research-based strategies that work*. New York: Guilford Press.
- Lane, K. L., Menzies, H., Kalberg, J. R., & Oakes, W. P. (2012). A comprehensive, integrated three-tier model to meet students' academic, behavioral, and social needs. In K. Harris, T. Urdan, & S. Graham (Eds.), *APA Educational psychology handbook* (Vol. 3, pp. 551-581). Washington, DC: American Psychological Association.
- Lane, K. L., Menzies, H. M., Oakes, W. P., & Kalberg, J. R. (2012). *Systematic screenings of behavior to support instruction: From preschool to high school*. New York: Guilford.
- Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575. doi:10.1111/j.1744-6570.1975.tb01393.x
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385. doi:10.1002/nur.20147

- Marion, S. F., & Pellegrino, J. W. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, 25, 47-57. doi:10.1111/j.1745-3992.2006.00078.x
- McLaughlin, M., & Talbert, J. (2003). *Reforming districts: How districts support school reform*. Seattle, WA: Center for the Study of Teaching and Policy.
- McCart, A., Sailor, W., Bezdek, J., & Satter, A. (2014c). A framework for inclusive educational delivery systems. *Inclusion*, 2, 252-264. doi:10.1352/2326-6988-2.4.252
- McLeskey, J., Waldron, N. L., Spooner, F., & Algozzine, B. (2014). *Handbook of research and practice for inclusive schools*. New York: Routledge.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds), *Test validity* (pp. 33-59). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York: Macmillan.
- Messick, S. (1995a). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14, 5-8. doi:10.1111/j.1745-3992.1995.tb00881.x
- Messick, S. (1995b). Validity of psychological assessment: Validation of inferences from persons' response and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749. doi:10.1037/0003-066X.50.9.741
- Mitchiner, M. (2014). *Post hoc study of a state selection process to predict state readiness to participate in schoolwide inclusive school reform*. (Doctoral dissertation, University of Kansas).

Morsbach Sweeney, H., Horner, R., Algozzine, B., Lane, K, Roger, B., . . . Sailor, W. (2013).

Schoolwide integrated framework for transformation fidelity implementation tool:

Implementation manual (version 1.4). Lawrence, KS: SWIFT Center.

National Professional Development Center on Autism Spectrum Disorder. (2011). *Autism*

Program Environment Rating Scale-Preschool/Elementary (APERS-PE). Chapel Hill,

NC: Author. Retrieved from

<http://vtautismproject.wikispaces.com/file/view/PE%20APERS%20March%202011.pdf>

Poggio, J. P., Glassnapp, D. R., Miller, D. M., Tollefson, N., & Burry, J. A. (1986). Strategies for

validating teacher certification tests. *Educational Measurement: Issues and Practices*, 5,

18-25. doi:10.1111/j.1745-3992.1986.tb00474.x

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's

being reported? Critique and recommendations. *Research in Nursing & Health*, 29, 489-

497. doi:10.1002/nur.20147

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content

validity: Conducting a content validity study in social work research. *Social Work*

Research, 27, 94-104. doi:10.1093/swr/27.2.94

Sailor, W. (2014). Advances in schoolwide inclusive school reform. *Remedial and Special*

Education, 36, 94-99. doi:10.1177/0741932514555021

Sailor, W. (2009a). Access to the general curriculum: Systems change or tinker some more?

Research and Practice for Persons with Severe Disabilities, 33, 249-257.

doi:10.2511/rpsd.33.4.249

Sailor, W. (2009b). *Making RTI work: How smart schools are reforming education through*

school-wide response-to-intervention. San Francisco: Jossey-Bass.

- Sailor, W., & Roger, B. (2003). *SAMAN: An instrument for the analysis of selected critical features of the Schoolwide Applications Model (SAM)* (1st ed.). Lawrence, KS: SAMSCHOOLS, LLC.
- Sailor, W., & Roger, B. (2005). Rethinking inclusion: Schoolwide applications. *Phi Delta Kappan*, 86(7), 503-509.
- Sailor, W., & Roger, B. (2008). *SAMAN: An instrument for the analysis of selected critical features of the Schoolwide Applications Model (SAM)* (2nd ed.). Lawrence, KS: SAMSCHOOLS, LLC.
- Sailor, W., Wolf, N., Choi, H., & Roger, B. (2009). Sustaining positive behavior support in a context of comprehensive school reform. In W. Sailor, G. Dunlap, G. Sugai & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 633-669). New York: Springer.
- Sailor, W., Zuna, N., Choi, J.-H., Thomas, J., McCart, A., & Roger, B. (2006). Anchoring schoolwide positive behavior support in structural school reform. *Research and Practice for Persons with Severe Disabilities*, 31, 18-30. doi:10.2511/rpsd.31.1.18
- Shogren, K. A., McCart, A. B., Lyon, K. J., & Sailor, W. S. (2015). All Means All Building Knowledge for Inclusive Schoolwide Transformation. *Research and Practice for Persons with Severe Disabilities*, 40, 173-191. doi:10.1177/1540796915586191
- Soukakou, E. P. (2012). Measuring the quality of inclusive classrooms: Development and validation of the *Inclusive Classroom Profile (ICP)*. *Early Childhood Research Quarterly*, 27, 478-488. doi:10.1016/j.ecresq.2011.12.003
- Sweetland, S. R., & Hoy, W. K. (2000). School characteristics and educational outcomes: Toward an organizational model of student achievement in middle schools. *Educational Administration Quarterly*, 36, 703-729. doi:10.1177/00131610021969173

- U.S. Department of Education. (2014). *35th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2013*. Washington, DC: Office of Special Education and Rehabilitative Services, Office of Special Education Programs.
- Waldron, N. L., & McLeskey, J. (2010). Establishing a collaborative school culture through comprehensive school reform. *Journal of Educational and Psychological Consultation*, *20*, 58-74. doi:10.1080/10474410903535364
- Ysseldyke, J. E., & Algozzine, B. (2006). *The legal foundations of special education*. Thousand Oaks, CA: Corwin Press.
- Ysseldyke, J. E., Algozzine, B., & Thurlow, M. (2000). *Critical issues in special and remedial education*. Boston, MA: Houghton Mifflin.

Table 1

SWIFT Fidelity of Implementation Tool Domain and Feature Level of Importance as Rated by Expert Panel Members

Domain/Feature	<i>M</i>	<i>Range</i>	<i>SD</i>	<i>CVI</i>
Administrative Leadership	3.67	2.75-4.00	0.35	.93
Strong and Engaged Site Leadership	3.75	2.60-4.00	0.36	.96
Strong Educator Support System	3.48	2.00-4.00	0.60	.87
Multi-tiered System of Support	3.71	3.00-4.00	0.30	.96
Inclusive Academic Instruction	3.72	2.92-4.00	0.34	.97
Inclusive Behavior Instruction	3.69	3.00-4.00	0.33	.95
Integrated Educational Framework	3.85	3.38-4.00	0.22	.98
Fully Integrated Organizational Structure	3.83	3.25-4.00	0.24	1.00
Strong and Positive School Culture	3.86	3.00-4.00	0.26	.95
Family and Community Engagement	3.59	2.71-4.00	0.41	.94
Trusting Family Partnerships	3.75	2.71-4.00	0.33	1.00
Trusting Community Partnerships	3.21	1.00-4.00	0.87	.80
Inclusive Policy Structure and Practice	3.48	3.00-4.00	0.41	.95
Strong LEA/School Relationship	3.50	2.60-4.00	0.47	.92
LEA Policy Framework	3.49	3.00-4.00	0.39	.98

Note. LEA refers to local educational agency.

Table 2

SWIFT Fidelity of Implementation Tool Domain and Feature Knowledge Development Sites (KDS) Pilot and Final Comparison

Domain/Feature	KDS Pilot		KDS Final		<i>t</i> - statistic ^a	<i>d</i> ^b
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Administrative Leadership	83.33	13.18	86.11	20.55	-0.25	.21
Strong and Engaged Site Leadership	86.67	17.89	86.67	17.64	0.00	.00
Strong Educator Support System	77.78	18.59	85.19	25.66	-0.50	.40
Multi-tiered System of Support	54.58	24.66	50.00	8.49	0.30	.19
Inclusive Academic Instruction	72.22	22.31	65.74	15.30	0.45	.29
Inclusive Behavior Instruction	34.72	34.73	18.51	8.48	0.77	.66
Integrated Educational Framework	67.28	19.23	77.78	13.39	-0.84	.55
Fully Integrated Organizational Structure	73.02	14.65	86.11	12.73	-1.31	.89
Strong and Positive School Culture	54.17	53.36	69.44	17.35	-0.47	.29
Family and Community Engagement	70.99	20.74	58.73	36.37	0.66	.59
Trusting Family Partnerships	59.26	18.14	64.44	40.73	-0.28	.29
Trusting Community Partnerships	70.37	19.46	44.44	25.46	1.72	1.33
Inclusive Policy Structure and Practice	44.70	41.28	37.78	18.95	0.27	.17
Strong LEA/School Relationship	52.78	41.20	35.56	25.24	0.63	.42
LEA Policy Framework	46.67	48.07	40.00	17.64	0.23	.14
Total	64.41	21.56	57.94	15.69	0.46	.30

Note. LEA refers to local educational agency.

^aDifference between means is not statistically significant at $p < .05$ level. ^bEffect size (d) = $(M_{Pilot} - M_{Final}) / SD_{Pilot}$

Table 3

*SWIFT Fidelity of Implementation Tool Domain and Core Feature Cohort and Knowledge**Development Sites (KDS) Comparison*

Domain/Feature	Cohort		KDS		<i>t</i> - statistic	<i>d</i> ^b
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Administrative Leadership	55.21	20.14	86.11	20.55	-2.26 ^a	1.53
Strong and Engaged Site Leadership	51.67	20.70	86.67	17.64	-2.58 ^a	1.69
Strong Educator Support System	61.11	23.76	85.19	25.66	-1.47	1.01
Multi-tiered System of Support	21.99	10.16	50.00	8.49	-4.22 ^a	2.76
Inclusive Academic Instruction	19.79	7.63	65.74	15.30	6.88 ^a	6.02
Inclusive Behavior Instruction	26.39	22.37	18.51	8.48	0.58	0.35
Integrated Educational Framework	18.23	14.25	77.78	13.39	-6.25 ^a	4.18
Fully Integrated Organizational Structure	11.46	14.04	86.11	12.73	-8.01 ^a	5.32
Strong and Positive School Culture	25.00	16.67	69.44	17.35	-3.90 ^a	2.67
Family and Community Engagement	37.50	18.08	58.73	36.37	-1.34	1.17
Trusting Family Partnerships	30.00	19.52	64.44	40.73	-1.97	1.76
Trusting Community Partnerships	56.25	28.08	44.44	25.46	0.63	0.42
Inclusive Policy Structure and Practice	48.33	19.92	37.78	18.95	0.79	0.53
Strong LEA/School Relationship	40.00	22.82	35.56	25.24	0.28	0.19
LEA Policy Framework	56.67	19.84	40.00	17.64	1.27	0.84
Total	37.83	11.34	57.94	15.69	-2.32 ^a	1.77

Note. LEA refers to local educational agency.

^aDifference between means is statistically significant at $p < .05$ level. ^bEffect size (d) = $(M_{Pilot} - M_{Final}) / SD_{Pilot}$

Figure 1. SWIFT domains and core features.

Figure 2. SWIFT-FIT example: Items 5.1, 5.2. (Morsbach Sweeney et al., 2014, p. 20)

Figure 3. SWIFT-FIT scoring rubric. (Morsbach Sweeney et al., 2014, p. 33)