

SWIFT Center

SWIFT Fidelity of Implementation Tool: Development and Preliminary Technical Adequacy

Prepared by

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

June, 2014

SWIFT Fidelity of Implementation Tool: Development and Preliminary Technical Adequacy

Abstract

The SWIFT Fidelity of Implementation Tool (SWIFT-FIT) is designed to measure growth and maturation of a school's inclusive educational practices and to help simplify school decision making about installing or improving practice. This report describes the development and technical adequacy of this instrument. Based on this technical adequacy, trained assessors may use SWIFT-FIT in research, evaluation and on-going implementation of the SWIFT framework in schools.

Introduction

Schoolwide Integrated Framework for Transformation (SWIFT) is a whole-system reform model that addresses six critical issues facing America's 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 are supported by extensive professional writing (see Table 1).

SWIFT Fidelity of Implementation Tool (SWIFT-FIT) provides a measure of the extent to which school personnel are using inclusive educational practices that align with SWIFT domains and features. The purpose of this report is to document the development process and preliminary technical adequacy of this tool. Development involved a multi-step collaborative and iterative process. Evaluating preliminary technical adequacy involved psychometric analyses of a sample of completed measures as well as completion of series of studies designed to provide support for key features of SWIFT-FIT validity, reliability, and usability.

Domains and Features

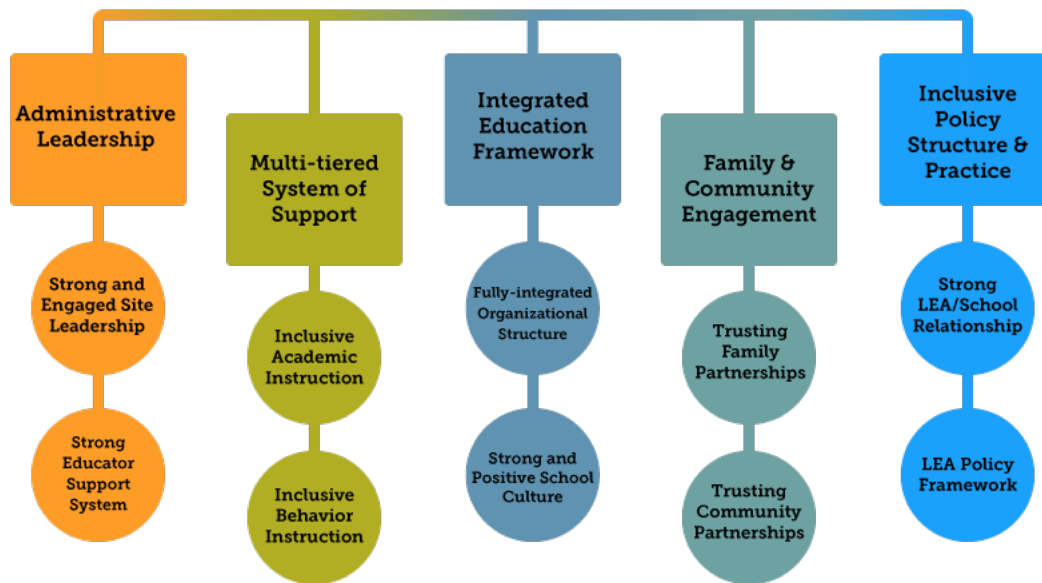


Figure 1. SWIFT domains and core features.

Developing SWIFT-FIT

We followed accepted and widely recommended stages for development of assessment instruments in counseling, education, psychology, and other social science areas (cf. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Dimitrov, 2012; Messick, 1989, 1995; 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 the development draft of the measure.

Purpose

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 (Sailor, 2012). Teams were guided by the following definition: Fidelity of implementation is the extent to which an intervention is delivered as conceived and planned. SWIFT envisioned a tool that would measure growth and maturation of a

school’s inclusive educational practices, while at the same time, help to simplify school decision making about installing or improving practice. Thus, SWIFT-FIT is intended for use in research, evaluation and on-going implementation of the SWIFT framework. School personnel may use SWIFT-FIT to assess “...the extent to which they already are using core features of SWIFT, to plan for training and technical assistance to improve their implementation of SWIFT, and to objectively document SWIFT adoption” (Morsbach Sweeney et al., 2013, p.5).

Table 1. SWIFT Domains, Descriptions, and Supportive Literature

SWIFT Domain	Description	Supportive Literature
Administrative Leadership	Strong and engaged site leadership committed to SWIFT processes, improving teaching and learning, and a system that empowers educators and school personnel.	Bedell & Burrello, 2006 Burrello, Hoffman, & Murray, 2005 Marzano, Waters, & McNulty, 2005 Vescio, Ross, & Adams, 2008 Waldron, McLeskey, & Redd, 2011
Multi-Tiered System of Support	Continuum of research-based, system wide practices to meet the identified academic and behavior instructional needs of students using data.	Bender & Shores, 2007 Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011 McCook, 2006 Sailor, 2009 Sailor & Roger, 2005 Shores & Chester, 2009 Waldron & McLeskey, 2010
Integrated Educational Framework	A framework that encompasses ALL students, personnel and stakeholders within a positive school culture and ensures full access for ALL students to participate in all school-related activities.	Farrell, Dyson, Polat, Hutcheson, & Gallannaugh, 2007a,b Hang & Rabren, 2009 Linn, 2011 McLeskey & Waldron, 2002, 2006, 2011 Sailor & Roger, 2005 Waldron & McLeskey, 2010 Waldron, McLeskey, & Redd, 2011 Wenger, 2000
Family and Community Engagement	The school, its families, and community members form trusting partnerships where each benefit from and support one another.	Anders-Butcher, Lawson, Bean, Flaspohler, Boone, & Kwiatkowski, 2008 Lawson & Sailor, 2000 McLeskey & Waldron, 1996, 2002, 2006, 2007, 2011 McLeskey, Waldron, & Redd, 2014 Sailor & Roger, 2005 Waldron, McLeskey, & Redd, 2011
Inclusive Policy Structure and Practice	Strong supportive relationships between school and its local educational agency through which school resource decisions become a matter of trust and respect and are supported within a policy framework that is fully aligned and actively supports the removal of barriers to the implementation of SWIFT.	Bryk, Bender-Sebring, Allensworth, Luppescu, & Easton, 2010 Fixsen, Blase, Metz, & VanDyke, 2010 Jerald, 2005 Kozleski & Smith, 2009 McLeskey, Waldron, & Redd, 2014 Sailor & Roger, 2005 Waldron & McLeskey, 2010 Waldron, McLeskey, & Redd, 2011

Technical Features

Technical features of SWIFT-FIT were identified by SWIFT development and review teams, which were led by persons with extensive experience in their respective fields and considered experts for the content they developed or reviewed. As the first step in developing SWIFT-FIT, the development team explored extant and supportive knowledge about the essential domains and features of SWIFT. 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 SWIFT feature to serve as base for documenting fidelity, and collected feedback from members of seven multi-organizational teams that comprise the SWIFT National Leadership Consortium (NLC).

All 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 of SWIFT-FIT items until consensus was reached.

Several versions of the SWIFT-FIT were developed to refine its content and organization during a pilot and assessor training period that took place at a school in Kansas City. The first version was ready in August, 2013 and contained total 53 items across the 10 core features. After the two pilots/assessor trainings in September, 2013, the total number of items was reduced to 51. The major changes were made to the Inclusive Academic Instruction core feature. SWIFT-FIT items related to research-based Tier I reading and math core curricula and instruction (i.e., practice) were separate in the first version, allowing measurement of reading Tier I core curricula, reading Tier I core instruction, math Tier I core curricula, and math Tier I core instruction. In the revised version, core instruction and instruction items were combined (i.e., research-based Tier I core curricula that is instructed with fidelity), but were still asked separately for reading and math. This revision reduced the number of items in Inclusive Academic Instruction from 13 to 11, which changed the total number of items from 53 to 51. Further, the item regarding culturally appropriate practice in the Inclusive Behavior Instruction feature was generalized and moved to the Strong and Positive School Culture feature. Score criteria were also refined based on the first two pilots/assessor trainings.

Between the second and third pilots/assessor trainings, data sources (i.e., possible interview, document review, and observation) were refined to capture better evidences to score each item. After the third pilot, two items in Inclusive Behavior Instruction were integrated to jointly measure the effective use of both outcome and fidelity data. This revision reduced the total number of items to 50.

In the final version of SWIFT-FIT, the number of items in Inclusive Academic Instruction changed from 11 to 12 because the item regarding academic progress monitoring was divided into progress monitoring for reading and math. The final version of the SWIFT-FIT, therefore, included 51 items. This process of literature review, expert analysis, and consensus established the best combined knowledge as a basis for SWIFT-FIT technical features.

Core Content

The final version of SWIFT-FIT includes 51 items across five domains and ten core features. The tool assesses variation in the extent to which the intervention is implemented as conceived and planned (i.e., fidelity) using item scores (i.e., ratings of 0, 1, 2, 3); feature scores (i.e., percentage of points for each core feature); domain scores (i.e., percentage of points for each domain); and a total score (i.e., average of the feature scores). No weighting of items is included in any SWIFT-FIT score calculation.

Preliminary Technical Adequacy

The evaluation team followed evidence-based practices for documenting the technical adequacy of measurement data (i.e., scores) obtained using assessment instruments in counseling, education, psychology, and other social science areas (cf. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Dimitrov, 2012; Messick, 1989, 1995; Rubio et al., 2003; Soukakou, 2012). They used data from several sources and documented selected features of the validity, reliability, and usability (see Table 2) as preliminary evidence of the technical adequacy of SWIFT-FIT.

Table 2. Documented Features of SWIFT Fidelity of Implementation Tool Technical Adequacy

Validity	Analysis	Guiding Question	Method
Content-Related	Face	To what extent does SWIFT-FIT “looks like” an acceptable and reasonable indicator of an inclusive educational framework?	Face Validity
	Expert	To what extent do expert panel members (e.g., Knowledge Development Site representatives) believe items and measure are adequate and appropriate indicators of levels of implementation?	Expert Validity
Construct-Related	Logical	To what extent do scores on SWIFT-FIT reflected naturally-occurring variation in implementation?	Logical Validity
	Concurrent	To what extent are scores on SWIFT-FIT related to scores on a similar measure administered at the same time?	Concurrent Validity
	Predictive	To what extent are scores on SWIFT-FIT related to scores for similar measure administered at a later time?	Predictive Validity
Reliability	Test-Retest	To what extent are scores on SWIFT-FIT similar for repeated assessments?	Test-Retest
	Internal Consistency	To what extent are scores on SWIFT-FIT similar across items or subsets of items?	Internal Consistency

Validity	Analysis	Guiding Question	Method
	Rater Agreement	To what extent are scores on SWIFT-FIT similar for different assessors?	Rater Agreement
Usability	Trained Assessor	To what extent do assessors favorably rate the ease of use and other procedural aspects of SWIFT-FIT?	Usability

Data Sources

The evaluation relied on three types of venues for data collection: Knowledge Development Sites (KDS), assessor training sites, and SWIFT partner sites. As a part of the larger SWIFT study, six schools were selected through a multi-step process as exemplary sites evidencing positive outcomes for students within inclusive settings. First, using a key informant strategy, members of the NLC—which is comprised of researchers and technical assistance providers representing key features of the SWIFT framework—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 achievement growth of students (including those with disabilities). An initial screening of these areas as well as additional grant requirements reduced the initial sample to 32 schools. Next, survey and phone interviews were conducted to gather additional information about the composition of these schools and more in-depth understanding of the ways in which they implemented key SWIFT features. The leadership team then reviewed this information and weighed considerations related to geographic location, population demographics, and ages served; they narrowed the sample to 11 schools. Finally, a team consisting of five SWIFT researchers and technical assistance providers visited schools. One-day visits included guided interviews and the completion of assessments of effective practices specific to SWIFT. Data were summarized, strengths and weaknesses considered, and ultimately six schools were selected as KDS where further study would take place. Serving as examples of effective inclusive schools, each KDS had evidence of some, but not all, core features in the SWIFT framework.

The second type of venue was Kansas City area schools, where trainers and assessors in training conducted SWIFT-FIT assessments. These assessor training sites were selected because someone on the SWIFT team had a previous relationship with the district, principal, or member of the school staff and obtained permission make observations in the school for SWIFT-FIT assessor training. From 22 requests, nine schools initially agreed to participate; however, one school opted out due to scheduling conflicts. The remaining eight schools participated as data sources in the SWIFT-FIT evaluation process, and are referred to here as Cohort sites.

The third type of venue was SWIFT partner schools. Personnel from each of SWIFT’s participating State Education Agency (SEA; Maryland, Oregon, Mississippi, Vermont and New Hampshire) selected two Local Educational Agencies (LEA) to identify

individual schools to serve as SWIFT partner sites. SWIFT partners participate in intensive technical assistance for installing and sustaining an inclusive educational framework. At the time of this technical adequacy study, Maryland, Oregon and Mississippi SEAs each chose 16 schools; Vermont SEA chose 11 schools (across 4 LEA) and; New Hampshire SEA chose 8 schools, for a total of 67 schools that met the following criterion. These schools are located in varying areas across each state, with no more than six sites in any one LEA. Partner sites are in urban (21%), suburban (6%) and rural (73%) communities and range in size (average student enrollment = 397, $SD = 182$, Range = 53-938) and demographics.

Validity

For content-related evidence, we documented the extent to which SWIFT-FIT scores “looked like” an acceptable and reasonable indicator of implementation (i.e., face validity; e.g., % of items that are similar to those on previous or similar assessment). We also documented expert validity, that is, the extent to which expert panel members (e.g., KDS representatives) 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 (i.e., logical validity); the extent to which scores on SWIFT-FIT were related to scores on a similar measure administered at the same time (i.e., concurrent validity); and the extent to which scores on SWIFT-FIT were related to scores for a similar measure administered at a later time (i.e., predictive validity).

Face validity. We documented the extent to which SWIFT-FIT “looks like” an acceptable and reasonable indicator of an inclusive educational framework by comparing its content and that of two other measures used to assess implementation levels of inclusive education with a previous model similar to SWIFT. Schoolwide Applications Model Analysis of Selected Critical Features (SAMAN) is a valid and reliable tool that was created to monitor implementation fidelity of the Schoolwide Applications Model (SAM) (Sailor & Roger, 2005). The SAMAN includes assessment items for each of the 15 critical features of SAM (Sailor & Roger, 2008). Each feature is scored using a 4-point rubric (see Appendix A). The second measure, the SAM Benchmark Fidelity Tool v1.2 (BFT) (Morsbach Sweeney, Rohrbaugh, McCart, Wolf, Sailor, & Roger, 2012) was created for use with SAM implementation in Washington, D.C. schools. BFT was created to assess the implementation of five key areas that serve as a strong foundation for SAM and includes 26 items across five key areas, which are scored using a 4-point rubric (see Appendix B). Reliability and validity of the BFT has not been examined.

SAMAN is the fidelity tool that was used for the framework of inclusive education from which the SWIFT framework emerged. Lessons learned from work with the SAM

model of inclusive education were later included into the SWIFT framework and subsequently into SWIFT-FIT for measurement. Content similarity with the SAMAN and BFT was 73.6% and 54.7%, respectively.

For SAMAN, more than 80% agreement was evident for content related to strong and engaged site leadership, inclusive academic and behavior instruction, fully integrated organizational structure, and family and community engagement. Additional areas of focus included in SWIFT-FIT, based on previous use of the SAMAN, included the importance of (a) 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.

Table 3. SWIFT Fidelity of Implementation Tool Content Similarity with Schoolwide Applications Model Analysis of Selected Critical Features (SAMAN) and Benchmark Fidelity Tool (BFT)

Domain/Feature	Content	Comparison scale	
		SAMAN	BFT
Domain 1	Administrative Leadership	62.5% (5/8)	37.5% (3/8)
Feature 1	Strong and Engaged Site Leadership	100%	40%
Feature 2	Strong Educator Support System	0%	33%
Domain 2	Multi-tiered System of Support	94.4% (17/18)	100% (18/18)
Feature 3	Inclusive Academic Instruction	100%	100%
Feature 4	Inclusive Behavior Instruction	83%	100%
Domain 3	Integrated Educational Framework	87.5% (7/8)	50% (4/8)
Feature 5	Fully Integrated Organizational Structure	100%	75%
Feature 6	Strong and Positive School Culture	75%	25%
Domain 4	Family and Community Engagement	100% (7/7)	57.1% (4/7)
Feature 7	Trusting Family Partnerships	100%	80%
Feature 8	Trusting Community Partnerships	100%	0%
Domain 5	Inclusive Policy Structure and Practice	30% (3/10)	0% (0/10)
Feature 9	Strong LEA/School Relationship	40%	0%
Feature 10	LEA Policy Framework	20%	0%
Total		73.5% (39/53)	54.7% (29/53)

Although SAMAN was designed for overall fidelity of implementation assessment, BFT was developed in response to the District of Columbia Public Schools (DCPS) schools implementation of SAM, primarily in four areas: leadership, multi-tiered system of support, family engagement, and a fully integrated organizational structure. Relationships between SWIFT-FIT and BFT were higher in these areas and, as would be expected, lower for content areas that were not part of the DCPS implementation (i.e., administrative leadership, trusting community partnerships, and inclusive policy structure).

Overall, SWIFT-FIT “looks like” an acceptable and reasonable indicator of an inclusive educational framework. Additional analyses supported its content-related validity.

Expert validity. Using an online survey, we documented the extent to which expert panel members rated SWIFT-FIT content as adequate and appropriate indicators of levels of implementation (see Appendix C). 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 results 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 *not very important* (1) to *very important* (4). 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 link to the survey to each individual participant and explained the reason they were chosen for participation, stated the purpose of the survey, and requested their participation. A week later, another email was sent thanking those who had participated and requesting those who had not yet to please take the survey. Results from the online software program, *Qualtrics*, indicated that the survey took on average 12-15 minutes to complete and had a 76% response rate (19 of 25 participants).

Expert ratings of importance across each SWIFT domain and core feature are summarized in Table 4. 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) (see Appendix C). The lower rating for universal screening for behavior is likely due to the lack of experience with and knowledge about the use of the practice. All KDS reported unfamiliarity with this practice and few used systematic tiered supports for behavior. They reported that high levels of academic engagement and added supports in each classroom provided proactive efforts, which resulted in “few behavior challenges.”

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: working collaboratively to monitor and plan academic interventions, and collaboratively teach all students in fully inclusive classrooms.

To obtain a statistical summary of expert validity, we randomly selected ratings from ten experts for further analysis. Representing the proportion of experts who rate the 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 (cf. Beck & Gable, 2001; Dimitrov, 2012; Polit & Beck, 2006; Rubio et al., 2003). 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 .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.

Table 4. 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

The average proportion of items that received ratings of important or very important on SWIFT-FIT was .94. The average CVI for domains of influence ranged from .93 (Administrative Leadership) to .98 (Integrated Educational Framework; see Table 4). The average CVI for features ranged from .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.

Logical 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

each of the six KDS and administered the final version in three of these schools approximately six month later. We also collected SWIFT-FIT scores in eight Cohort schools. 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 5. Domain, feature, and total scores were statistically similar. SWIFT-FIT total scores for the Pilot ($M = 64.41$, $SD = 21.56$) and Final ($M = 57.94$, $SD = 15.69$) versions were not statistically significantly different ($t = 0.46$, $d = .30$). Effect sizes were small ($d = .17-.21$) for Administrative Leadership, Multi-tiered System of Support, and Inclusive Policy Structure and Practice; and moderate for Integrated Educational Framework ($d = .55$) and Family and Community Engagement ($d = .59$) score differences. Similarly, small practical differences ($d = .00-.42$) were evident for all features except Inclusive Behavior Instruction ($d = .66$), Fully Integrated Organizational Structure ($d = .89$), and Trusting Community Partnerships ($d = 1.33$). In general, the observed similarities were greater than differences across the two versions of SWIFT-FIT (see Figure 2).

Means, standard deviations, and comparison statistics across Cohort and KDS schools are in Table 6. SWIFT-FIT total scores for Cohort ($M = 37.83$, $SD = 11.34$) and KDS ($M = 57.94$, $SD = 15.69$) schools were statistically significantly different ($t = -2.32$, $d = 1.77$). Statistically significant differences and large effect sizes were also evident for Administrative Leadership ($t = -2.26$, $d = 1.53$), Multi-tiered System of Support ($t = -4.22$, $d = 2.76$), and Integrated Educational Framework ($t = -6.25$, $d = 4.18$) domain scores; however, Family and Community Engagement ($M = 37.50$, $SD = 18.08$; $M = 58.73$, $SD = 36.37$) and Inclusive Policy Structure and Practice ($M = 48.33$, $SD = 19.92$; $M = 37.78$, $SD = 18.95$) domain scores were statistically similar ($t = -1.34$, $d = 1.17$; $t = 0.79$, $d = .53$) for Cohort schools and KDS, respectively. Large practical differences ($d = 1.69-6.02$) were evident for four core features including Strong and Engaged Site Leadership ($d = 1.69$), Inclusive Academic Instruction ($d = 6.02$), Fully Integrated Organizational Structure ($d = 5.32$), and Strong and Positive School Climate ($d = 2.67$); non-statistically significant, but large practical differences were evident for Strong Educator Support System ($d = 1.01$), Trusting Family Partnerships ($d = 1.76$), and LEA Policy Framework ($d = 0.84$); and non-statistically significant and small practical differences were evident for Inclusive Behavior Instruction ($d = 0.35$), Trusting Community Partnerships ($d = .42$), and Strong LEA/School Relationship ($d = 0.19$). Observed scores were higher for KDS than Cohort schools (see Figure 3).

Table 5. SWIFT Fidelity of Implementation Tool Domain and Feature Knowledge Development Sites (KDS) Pilot and Final Comparison

Domain/Feature	KDS Pilot		KDS Final		t-statistic ¹	d ²
	M	SD	M	SD		
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

¹Difference between means is not statistically significant at $p < .05$ level.

²Effect size (d) = $M_{Pilot}/M_{Final}/SD_{Pilot}$

Table 6. SWIFT Fidelity of Implementation Tool Domain and Core Feature Cohort and Knowledge Development Sites (KDS) Comparison

Domain/Feature	Cohort		KDS		t-statistic	d ²
	M	SD	M	SD		
Administrative Leadership	55.21	20.14	86.11	20.55	-2.26 ¹	1.53
Strong and Engaged Site Leadership	51.67	20.70	86.67	17.64	-2.58 ¹	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 ¹	2.76
Inclusive Academic Instruction	19.79	7.63	65.74	15.30	-6.88 ¹	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 ¹	4.18
Fully Integrated Organizational Structure	11.46	14.04	86.11	12.73	-8.01 ¹	5.32
Strong and Positive School Culture	25.00	16.67	69.44	17.35	-3.90 ¹	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 ¹	1.77

¹Difference between means is statistically significant at $p < .05$ level.

²Effect size (d) = $M_{Pilot}/M_{Final}/SD_{Pilot}$

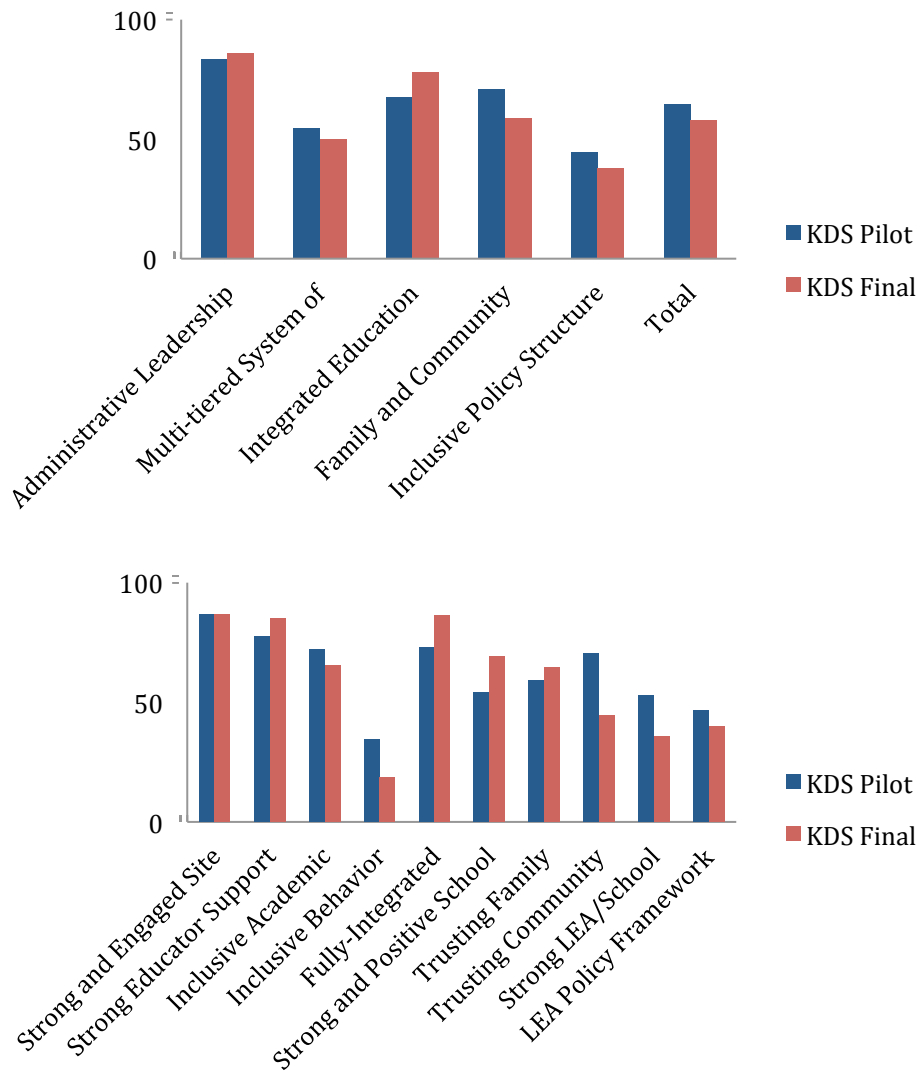


Figure 2. Total, domain and feature score comparisons across pilot and final versions of SWIFT Fidelity of Implementation Tool for Knowledge Development Sites (KDS).

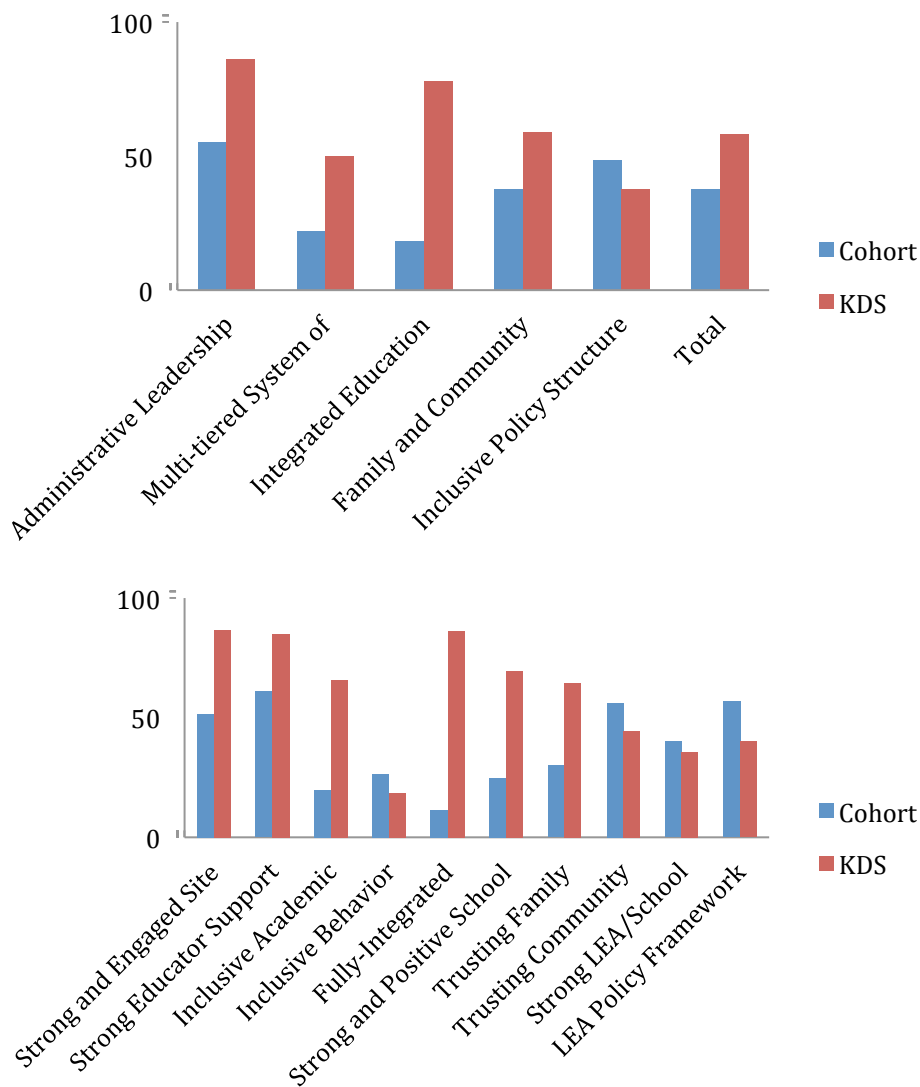


Figure 3. Total, domain and feature score comparisons across different levels of implementation represented by Knowledge Development Sites (KDS) with independently implemented inclusive practices and Cohort schools with no explicitly implemented inclusive practices.

We expected that scores for the KDS, which are recognized as models of inclusive educational practices, would be high and they were, suggesting that domain and core features “reflect” or “represent” reasonable implementation indicators. We also expected to document small effect sizes between the two SWIFT-FIT administrations and we did for 3 of the 5 domains and 7 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. Inclusive Behavior Instruction ($d = .66$) 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 ($d = .89$) had significant changes between administrations with 6 of its 8 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 the 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 closely matched their existing practices. The feature of Trusting Community Partnership ($d = 1.33$) 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.

As we expected, scores in KDS were higher than those in non-implementing Cohort schools. We expected to see large effect sizes for differences in scores for Cohort and KDS schools, and for 4 of the 5 domains and 7 of the 10 features, we did. Three features showed small effect sizes indicating no large differences between the two groups' scores: Inclusive Behavior instruction ($d = 0.35$), Trusting Community Partnerships ($d = .42$), and Strong LEA/School Relationship ($d = 0.19$). These findings follow the pattern evident in SWIFT-FIT scores for KDS related to Inclusive Behavior Instruction and Trusting Community Partnerships, however the results for the Inclusive Policy Structure and Practice domain, which includes the Strong LEA/School Relationship feature are different. Although one of the schools, for which data was included in this analysis, first had the full support of their district they reported that in recent years, the focus of the district had shifted and they were "lone wolves" now, which may explain lower scores in these areas. Another possible explanation for the results for this domain favoring the Cohort schools over the KDS is that many of the items in this domain are written specifically to measure SWIFT district level expectations and did not necessarily "fit" the Cohort schools. The questions conducted during interviews with Cohort schools, as a result, were altered in such a way for better understanding of those being interviewed and to better fit the context of non-inclusive schools, possibly resulting in higher scores.

Concurrent validity. To document the extent to which scores are related to those for similar or different measure 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 the two measures was similar enough to use the evidence gathered from the schools for SWIFT-FIT to concurrently score the SAMAN. The administration protocols for each require classroom observation, interviews with stakeholders, and document review. SWIFT-FIT administration requires not only additional depth but also breadth of evidence to be collected and as such was sufficient for scoring the SAMAN. The results from our face validity study comparing 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, core feature and total scores. Statistically significant correlations were evident for SWIFT-FIT total score (.70) and the 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 the Inclusive Policy Structure and Practice domain and the total score on the SAMAN. Statistically significant correlations were also evident for five critical 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 Strong Educator Support System (.55) and Trusting Community Partnerships (.42); low relationships were evident for Inclusive Behavior Instruction (.19) and LEA Policy Framework (.03); and, a moderate negative relationship was evident for Strong LEA/School Relationship (-.33) critical features.

The strongest associations would be expected in areas that were similar between the SAM and SWIFT models, 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. The low or negative correlations found in the 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. Recall, the SWIFT framework evolved from a whole-school approach to inclusion known as SAM, and the SAMAN, which was used for this analysis, is the associated fidelity instrument for SAM. 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 the SAM framework.

Table 7. Item-Total Correlations and Internal Consistency Estimates for SWIFT Fidelity of Implementation Tool.

Domain	Feature	Item	r_{ss}	r_{tot}	α				
Administrative Leadership	Strong and Engaged Site Leadership	1.1	.81	.71	.88				
		1.2	.59	.50					
		1.3	.78	.75					
		1.4	.68	.58					
		1.5	.80	.66					
	Strong Educator Support System	2.1	.83	.81		.76			
		2.2	.85	.74					
		2.3	.78	.56					
		<hr/>							
Multi-tiered System of Support	Inclusive Academic Instruction	3.1	.72	.67	.92				
		3.2	.49	.42					
		3.3	.69	.59					
		3.4	.71	.63					
		3.5	.90	.81					
		3.6	.84	.73					
		3.7	.70	.58					
		3.8	.59	.56					
		3.9	.91	.80					
		3.10	.78	.60					
		3.11	.79	.80					
		3.12	.40	.52					
	Inclusive Behavior Instruction	4.1	.84	.47		.84			
		4.2	.37	.24					
		4.3	.82	.52					
		4.4	.79	.57					
		4.5	.71	.31					
		4.6	.85	.56					
	Integrated Educational Framework	Fully Integrated Organizational Structure	5.1	.70		.23	.69		
			5.2	.58		.33			
			5.3	.42		.61			
5.4			.76	.63					
Strong and Positive School Culture		6.1	.80	.65	.68				
		6.2	.69	.44					
		6.3	.68	.47					
		6.4	.69	.39					
		<hr/>							
		Family and Community Engagement	Trusting Family Partnerships	7.1		.67		.46	.84
7.2	.75			.56					
7.3	.77			.49					
7.4	.77			.80					
7.5	.84			.77					
Trusting Community Partnerships	8.1		.79	.56	.47				
	8.2		.83	.50					
	<hr/>								
	Inclusive Policy Structure and Practice		Strong LEA/School Relationship	9.1		.63	.17	.84	
				9.2		.86	.59		
				9.3		.62	.43		
9.4		.59		.41					
9.5		.62		.51					
LEA Policy Framework		10.1	.89	.59	.79				
		10.2	.80	.76					
		10.3	.71	.38					
		10.4	.64	.45					
		10.5	.70	.55					
		<hr/>							
		Total							.96

Note. r_{ss} = correlation between the SWIFT-FIT item and the corresponding subscale. r_{tot} = correlation between the SWIFT-FIT item and the total mean score. α = Cronbach's alpha as an indicator of internal consistency. $n = 64$ partner schools

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. An internal consistency analysis was conducted to determine whether items in a domain and core features are homogeneous. This shows inter-relatedness of the items within the SWIFT-FIT to describe the extent to which all items measure the same concept or construct. To document internal consistency reliability estimates we calculated Cronbach's alpha (α) (Cronbach, 1969; Cronbach & Meehl, 1955) across partner site schools for five domains, ten core features, and all 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_{ss}) and between each item the total SWIFT-FIT score (r_{tot}) were also calculated (cf. Cohen et al., 2007).

In general, the reliability of SWIFT-FIT scores was high and above the general cutoff of interest for research purposes (Henson, 2001). Cronbach's alphas for Total SWIFT-FIT and the Core Features of 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. The Cronbach's alphas of core features ranged from .44 to .92 (see Table 7). 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 are also reported in Table 7. These analyses 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 (items 9.1, 4.2, 5.1, 4.5, 5.2, 10.3, and 6.4).

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. The schools included both practice schools and partner schools. Practice schools were those who had agreed to be used for the purpose of training assessors

and not part of the SWIFT Center technical assistance partnerships. SWIFT partner schools were those working with the SWIFT Center to implement the SWIFT framework.

For the 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 scored the SWIFT-FIT on a second copy of the tool. Observers were SWIFT-FIT assessors in training and had received preliminary training on the SWIFT domains and features as well as SWIFT-FIT administration protocols. In some cases, the observers were conducting their first session with a trained assessor, and in some it was 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), interobserver agreement was calculated between each pair, resulting in 30 pairings for analysis. We report the interobserver agreement between trained assessors and assessors in training to provide preliminary evidence as to the reliability of the tool.

The average interobserver agreement on SWIFT-FIT items across the 13 schools and 30 possible pairings of assessors was 79.61% (Range = 60.0%-96.08%). Excluding the interobserver agreement calculations between two observers (on those occasions where there was a primary assessor and two observers) the average was 82.29% (Range = 62.0%-96.08%).

Usability

To document the extent to which assessors favorably rated the SWIFT-FIT ease of use, an online survey was designed and administered using *Qualtrics* (see Appendix D). Participants rated each of 12 statements using a 4-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (4). Each of the trained assessors had administered SWIFT-FIT a minimum of twice. The 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 link to the survey was emailed to each of the participants to individually explain the reason they were chosen for participation, the purpose of the survey and asking for

their participation. A week later, another email was sent thanking those who had participated and requesting those who had not yet 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.

Means and standard deviations for respondents' ratings of usability for SWIFT-FIT are presented in Table 8. 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 the SWIFT-FIT assessment; assessor notes which are to be used to clarify intent of items on SWIFT-FIT and a listing of positions to interview and 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 ($M = 2.89$), SWIFT-FIT as an effective or efficient assessment of implementation fidelity ($M = 2.67$ and 2.89 , respectively). Overall usability of SWIFT-FIT was had a mean of 2.56.

Table 8. SWIFT Fidelity of Implementation Tool Usability Results as Indicated by Trained Assessors

Survey Item	<i>M</i>	<i>SD</i>
1 The purpose of SWIFT-FIT is described clearly in the protocol.	3.67	0.35
2 The steps in a SWIFT-FIT administration are described clearly in the protocol.	3.11	0.60
3 The process for preparing schools for a SWIFT-FIT administration is described clearly in the protocol	3.00	0.50
4 The process for scoring and using results from a SWIFT-FIT administration is described clearly in the protocol.	3.11	0.60
5 The item descriptions were clearly written and easy to understand.	3.00	0.50
6 The assessor notes were helpful in clarifying the intent of the items.	3.44	0.53
7 The positions to interview and questions sections were helpful in conducting the assessment.	3.44	0.53
8 The scoring rubrics were clearly written and easy to use.	2.89	0.33
9 In a training manual, additional information to help determine how to rate each item would be helpful.	3.67	0.71
10 The process for completing SWIFT-FIT is reasonable, appropriate, and effective.	2.89	0.33
11 SWIFT-FIT is an effective assessment of implementation fidelity effective.	2.67	0.50
12 SWIFT-FIT is an efficient assessment of implementation fidelity.	2.89	0.33

13	How would you rate the overall usability of SWIFT-FIT?	2.56	0.53
----	--	------	------

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. Most of these improvements were identified in the comments provided by respondents to the open ended questions, “What do you like most / least about the SWIFT-FIT?” and “What, if any improvements would you suggest for the SWIFT-FIT?” and in anecdotal information received from the assessors. Respondents’ comments on the survey indicated agreement about what they liked best about SWIFT-FIT, such as it is comprehensive, allows one to “really look into or understand the school” and “provides an overall snapshot of the implementation of a huge/complex framework.” Although these were recognized as strengths, they also conversely appeared to serve as what respondents liked least about SWIFT-FIT. Comments provided on the survey included that it is “very time consuming,” “feels way too big,” “amount to cover,” and “hard to navigate.” These perceptions were also evident in the ratings for the process of completing SWIFT-FIT being efficient, the tool itself viewed as reasonable and its overall usability, each of which had means below 3.0 or *agreed*. Improvements suggested on the survey along with anecdotal information received from the assessors fell into three main categories: (a) altering content of one of the specific features, (b) comments related to how SWIFT-FIT results will be used to guide future administrations, and (c) desires for the SWIFT-FIT to provide qualitative evidence that it is not intended or designed to provide. Assessors suggested that improvements to the Strong and Positive School Culture feature be considered to include more items important to school climate/culture. Many people, as they become familiar with the SWIFT framework and the items on SWIFT-FIT that represent this feature, agree that it does not represent the traditional view of school culture.

These considerations are noted for version 2.0 of SWIFT-FIT. At this point, to better understand the second and third areas of improvement centered on how SWIFT-FIT results will be used to guide future administrations and the desire for it to provide more qualitative evidence, it is important to know that many of the SWIFT-FIT assessors responding to this survey serve in the additional role of technical assistance providers. SWIFT-FIT, due to its comprehensive nature of such a large framework, was designed to assess the *presence* of systems and/or practices that are likely (as identified by research) to produce positive outcomes. The SWIFT-FIT was not designed, nor would it be possible in one day to assess, the *quality* by which those systems and/or practices are implemented or the outcomes they produce. For example, SWIFT-FIT assesses whether or not school leadership teams use data on a monthly basis, but whether that data is then used toward enacting improvement is not assessed. As a technical assistance provider, one is often interested in the outcomes of a practice or system being used rather than simply whether or not a

system is in place. This perspective was evident in comments about what several reviewers wished that SWIFT-FIT would be able to do. As a result, many of these same assessors recognized the need for technical assistance when making decisions based on additional data in concert with SWIFT-FIT—such technical assistance is the mission of SWIFT Center.

Five Things We Know about SWIFT-FIT Preliminary Technical Adequacy

SWIFT-FIT is used to measure of the extent to which schools engage in inclusive educational practices in the SWIFT framework. Our preliminary analysis of the technical adequacy of the SWIFT-FIT supports the following conclusions:

1. Content of SWIFT-FIT adequately reflects the “big ideas” of inclusive practices.
2. Scores on SWIFT-FIT provide a basis for documenting variation in levels of implementation of inclusive practices.
3. Scores on SWIFT-FIT are comparable to scores on similar measures of implementation of inclusive practices.
4. Scores on SWIFT-FIT are similar for different assessors.
5. Assessors provide favorable ratings of the usability of SWIFT-FIT.

Suggested Citation

Algozzine, B., Morsbach Sweeney, H., Choi, H., Horner, R., Sailor, W., McCart, A., Satter, A., & Lane, K. (2014). *SWIFT Fidelity of Implementation Tool (SWIFT-FIT): Development and Preliminary Technical Adequacy*. Lawrence, KS: National Center on Schoolwide Inclusive School Reform: The SWIFT Center.

References

- 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.
- Anderson-Butcher, D., Lawson, H. A., Bean, J., Flaspohler, P., Boone, B., & Kwiatkowski, A. (2008). Community collaboration to improve schools: Introducing a new model from Ohio. *Children and Schools, 30*, 161-172. doi: 10.1093/cs/30.3.161
- Beck, C. T., & Gable, R. K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement, 9*(2), 201-215.

- Bedell, C. P., & Burrello, L. C. (2006). A cultural shift toward distributed leadership. *Journal of School Leadership, 16*(6), 740-771.
- Bender, W. N., & Shores, C. (2007). *Response to intervention. A practical guide for every teacher*. Thousand Oaks, CA: Corwin Press.
- Bryk, A., Bender-Sebring, P., Allensworth, E., Luppescu, S., and Easton, J. (2010) *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Burrello, L. C., Hoffman, L. P., & Murray, L. E. (2005). *School leaders building capacity from within*. Thousand Oaks, CA: Corwin Press.
- 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.
- 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., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Dimitrov, D. M. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405-432. doi: 10.1111/j.1467-8624.2010.01564.x
- Farrell, P., Dyson, A., Polat, F., Hutcheson, G., & Gallannaugh, F. (2007a). Inclusion and achievement in mainstream schools. *European Journal of Special Needs Education, 22*, 131-145. doi: 10.1080/08856250701267808
- Farrell, P., Dyson, A., Polat, F., Hutcheson, G., & Gallannaugh, F. (2007b). The relationship between inclusion and academic achievement in English mainstream schools. *School Effectiveness and School Improvement, 18*, 335-352. doi: 10.1080/09243450701442746
- Fixsen, D. L., Blase, K., Metz, A., & Van Dyke, M. (2010). *Implementation teams: The active interface between interventions and effective human services*. Chapel Hill, NC: FPG Child Development Institute, University of North Carolina.
- Guion, R. M. (1977). Content validity—The source of my discontent. *Applied Psychological Measurement, 1*, 1-10. doi: 10.1177/014662167700100103
- Hang, Q., & Rabren, K. (2009). An examination of co-teaching: Perspectives and efficacy indicators. *Remedial and Special Education, 30*, 259-268. doi: 10.1177/0741932508321018
- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.

- Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G. & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A Research Instrument for Assessing School-Wide Positive Behavior Support. *Journal of Positive Behavior Interventions, 6*, 3-12. doi: 10.1177/10983007040060010201
- Jerald, C. (2005). The implementation trap: Helping schools overcome barriers to change. *Policy brief* (pp. 1-12). Washington, DC: The Center for Comprehensive School Reform and Improvement.
- Kozleski, E. B., & Smith, A. (2009). The role of policy and systems change in creating equity for students with disabilities in urban schools. *Urban Education, 44*, 427-451. doi: 10.1177/0042085909337595
- Lawson, H., & Sailor, W. (2000). Integrating services, collaborating, and developing connections with schools. *Focus on Exceptional Children, 33*(2), 1-22.
- Linn, M. (2011). Inclusion in two languages: Special education in Portugal and the United States. *Phi Delta Kappan, 92*(8), 58-60.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*, 382-385. doi: 10.1002/nur.20147
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: ASCD.
- McCook, J. E. (2006). *The RTI guide: Developing and implementing a model in your schools*. Horsham, PA: LRP Publications.
- McLeskey, J., & Waldron, N. L. (1996). Responses to questions teachers and administrators frequently ask about inclusive school programs. *Phi Delta Kappan, 78*(2), 150.
- McLeskey, J., & Waldron, N. L. (2002). School change and inclusive schools: Lessons learned from practice. *Phi Delta Kappan, 84*(1), 65-72.
- McLeskey, J., & Waldron, N. L. (2006). Comprehensive school reform and inclusive schools. *Theory into Practice, 45*, 269-278. doi: 10.1207/s15430421tip4503_9
- McLeskey, J., & Waldron, N. L. (2007). Making differences ordinary in inclusive classrooms. *Intervention in School & Clinic, 42*, 162-168. doi: 10.1177/10534512070420030501
- McLeskey, J., & Waldron, N. L. (2011). Educational programs for elementary students with learning disabilities: Can they be both effective and inclusive? *Learning Disabilities Research & Practice, 26*, 48-57. doi: 10.1111/j.1540-5826.2010.00324.x
- McLeskey, J., Waldron, N. L., & Redd, L. (2014). A case study of a highly effective, inclusive elementary school. *The Journal of Special Education, 48*, 59-70. doi: 10.1177/0022466912440455
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). New York, NY: Macmillan.
- Messick, S. (1995). 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

- Morsbach Sweeney, H., Horner, R., Algozzine, B., Lane, K, Roger, B., Choi, H., McCart, A., & Sailor, W. (2013). *Schoolwide Integrated Framework for Transformation Fidelity Implementation Tool: Implementation Manual* (version 1.4). Lawrence, KS: National Center on Schoolwide Inclusive Education: The SWIFT Center.
- Morsbach Sweeney, H., Rohrbaugh, L., McCart, A., Wolf, N., Sailor, W., & Roger, B. (2012). *SAM Benchmark Fidelity Tool* (version 1.2). Lawrence, KS: Authors.
- 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. (2009). *Making RTI work*. San Francisco: Jossey-Bass.
- Sailor, W. (2012). *Schoolwide Integrated Framework for Transformation Center (SWIFT)*. ED Grant Number H326Y120005. Washington, DC: U.S. Department of Education, Office of Special Education and Rehabilitative Services.
- 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.
- Shores, C., & Chester, K. (2009). *Using RTI for school improvement: Raising every student's achievement scores*. Thousand Oaks, CA: Corwin Press.
- 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
- Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education, 24*, 80-91. doi: 10.1016/j.tate.2007.01.004
- Waldron, N. L., & McLeskey, J. (2010). Establishing a collaborative school culture through comprehensive school reform. *Journal of Educational & Psychological Consultation, 20*, 58-74. doi: 10.1080/10474410903535364
- Waldron, N. L., McLeskey, J., & Redd, L. (2011). Setting the direction: The role of the principal in developing an effective, inclusive school. *Journal of Special Education Leadership, 24*, 51-60.
- Wenger, E. (2000). Communities of practice and social learning systems. *Organization, 7*, 225-246. doi: 10.1177/135050840072002



The authors produced this report under U.S. Department of Education, Office of Special Education Programs Grant No. H325Y120005, National Center on Schoolwide Inclusive School Reform: The 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. While permission to reprint this publication is not necessary, the citation should be: Algozzine, B., Morsbach Sweeney, H., Choi, H., Horner, R., Sailor, S., McCart, A., Satter, A., & Lane, K. (2014). *SWIFT Fidelity of Implementation Tool (SWIFT-FIT): Development and Preliminary Technical Adequacy*. Lawrence, KS: National Center on Schoolwide Inclusive School Reform: The SWIFT Center. Appendices referenced in this report are available on request from SWIFT Center (swift@ku.edu).