Evaluation of implementation outcomes for the IMPACT implementation support platform

Melissa E DeRosier*, Thea Cox, and Edith Lindley

3C Institute, 4364 S. Alston Avenue, Suite 300, Durham, NC 27713, USA

Abstract

Background: The IMPACT implementation support platform is a measurement and feedback system specifically designed to scale evidence-based programs and practices (EBPPs) and support high-quality implementation. The purpose of this study was to evaluate use of the IMPACT software system in terms of its acceptability, appropriateness, feasibility, and likelihood of adoption.

Methods: Seventy-eight school-based providers (across 49 schools) participated in this study. Demographic and background information was collected in the fall of 2019. Thereafter, providers delivered the Social Skills Group Intervention (S.S.GRIN) EBPP with students as usual and used IMPACT to enter and track progress and process data for their S.S.GRIN groups. In the spring of 2020, providers completed a series of ratings to evaluate their experience with IMPACT. Providers' usage of IMPACT during the study was also tracked.

Results: Providers' ratings of IMPACT for each implementation outcome were significantly higher than average ($p < 0.001$), with Satisfaction, Superior Innovation, and Interface Quality being especially positive. There was a significant interaction effect where ratings were higher for providers who reported disliking implementation data tracking for S.S.GRIN at the start of the study. Greater usage was significantly associated with higher ratings of providers' capacity for tracking implementation data and a higher likelihood of recommending IMPACT to others. No significant differences in the patterns of results were found for demographic subgroups.

Conclusion: IMPACT was seen as acceptable, feasible, and appropriate by school-based S.S.GRIN providers who reported likely continued use (adoption) of the innovation. IMPACT was particularly well received by those who started the study with negative impressions of implementation data tracking for S.S.GRIN. This study supports the potential utility and value for supporting ongoing implementation of school-based programs. Future research is needed to evaluate IMPACT for other EBPPs and in other service delivery settings to determine the generalizability of this study's findings.

Background

Too often, behavioral health treatment and prevention programs shown to be effective under controlled research conditions show null or small effects when used in real-world service settings [1-4]. To realize population-based positive impacts, high-quality implementation must be sustained when EBPPs are moved to scale [5,6]. Over the past two decades, the science of implementation has identified a number of factors that can effectively narrow this research-to-practice gap [7-9]. The challenge now is to apply what we know about the supports needed to take EBPPs to scale to the development of functional systems that can be used feasibly and effectively for everyday practice [10-12]. As underscored by the Society for Prevention Research (SPR) Mapping Advances in Prevention Science (MAPS) IV Translation Research Task Force, building systemic capacity for EBPP scale-up is a primary "action step to move the needle" on population-level well-being [13].

Impact: IMPACT is a measurement and feedback system designed to scale EBPPs. While a variety of data may be measured in such a system, IMPACT focuses on data collection, analysis, and reporting of process and progress metrics. This focus is grounded in the measurement-based care (MBC) research literature showing how consistent measurement and use of data in these domains is particularly influential for achieving target outcomes [14-17]. Studies across an array of behavioral health interventions have found regular use of process and progress metrics is associated with more positive service outcomes (e.g., more client-centered, greater efficiency of care) and clinical outcomes (e.g., lower symptomatology, improved function; [18,19].

In terms of measuring process, IMPACT focuses on three areas drawn from Durlak and DuPre's taxonomy (2008) of factors that impact implementation, specifically (a) *fidelity* (adherence to the original program's dosage and treatment model), (b) *quality* (provider's skill and efficacy in delivering content), and (c) *responsiveness* (participant engagement in the program). These areas have been found to directly influence outcomes and are most likely to attenuate when an EBPP is moved to scale [20-22].

In terms of measuring progress, IMPACT enables tracking of both *intermediate outcomes* during the course of treatment and change in target *clinical outcomes* as a function of participation in the EBPP. Intermediate outcomes are those that predict ultimate outcomes of treatment and may be closely tied to intervention content. For example, for an EBPP designed to reduce clinical depression, intermediate outcomes could include assessment of optimism and stress as well as knowledge gains for content presented during treatment sessions. Typically, assessment of clinical outcomes utilizes established psychometrically sound instruments administered at specific time

points (e.g., pre-treatment, post-treatment, and 3-month follow-up), but may also include 'home-grown' measures of expected changes over the course of treatment (e.g., school absenteeism, interpersonal relations).

Despite evidence showing MBC enhances outcomes, measurement and use of data in everyday service settings are often missing or are inconsistent at best [23-25]. Barriers that contribute to this gap include cumbersome, complex systems that are difficult or frustrating to use; incompatibility with typical workflow; and the disconnect between entry of data and actionable results. During the software design process, we strove to address these barriers and thereby facilitate MBC for EBPPs utilizing IMPACT. Given that EBPP providers most often bear the burden of entering process and progress data, we approached software design with a primary focus on the provider as the end user. Our guiding philosophy was "ask only what you truly need to ask, minimize the need for duplicative data entry, and structure data entry forms and processes to be as simple and user friendly as possible." IMPACT's user interface (UI), navigation, and features were iteratively refined based on input from providers of a variety of EBPPs in school and community service settings.

Contextually relevant, timely, and actionable feedback are essential to support data-informed care [26-27]. If it is weeks or months before results are shared, that data are neither meaningful nor actionable. However, if providers can see, in real time, how their fidelity to the model is linked with participant progress, for example, the connection between data and practice is much more evident. Further, software such as IMPACT is able to automate personalized feedback for continuous quality improvement (CQI). For example, if fidelity drops below a predefined benchmark, the software alerts the provider and directs them to resources to support high-fidelity implementation (e.g., demo video to help prepare for the next session). The combination (and linking) of data-driven alerts, feedback, and recommendations with easy access to online training, coaching, and implementation (e.g., program materials) resources is intended to associate provider actions with participant progress and thereby increase motivation to use data to improve care.

S.S.GRIN: For the purposes of this evaluation study, we investigated use of IMPACT for providers of S.S.GRIN (*Social Skills Group Intervention*; [28]). S.S.GRIN is an in-person small group intervention with developmental versions for Pre-K through 5th grade students. S.S.GRIN is a structured manualized program that combines social learning and cognitive-behavioral techniques to build children's social skills (e.g., communication, cooperation, social initiation). Clinical trials have supported S.S.GRIN's efficacy for enhancing children's peer relationships (e.g., increased peer acceptance) and school-based adjustment (e.g., lower social anxiety and declines in aggressive behavior problems) compared to controls, both at immediate and one-year follow-up [29]. S.S.GRIN is currently implemented throughout the U.S. with thousands of students each year.

We created the custom IMPACT application by translating S.S.GRIN's existing implementation methods, data collection forms (attendance, fidelity checklists, participant progress monitoring ratings), and reporting specifications into the software platform. Figure 1 provides an example of a data entry form for providers to assess their fidelity for an S.S.GRIN session.

Aim

This study's primary purpose was to determine providers' perceptions of the IMPACT platform following use of the software to track S.S.GRIN implementation in schools. Our study assessed IMPACT in terms of four outcomes identified in Proctor and colleagues' implementation outcomes framework, specifically acceptability, appropriateness, feasibility, and adoption [1]. A secondary purpose was to examine the degree to which providers accessed and used IMPACT over the course
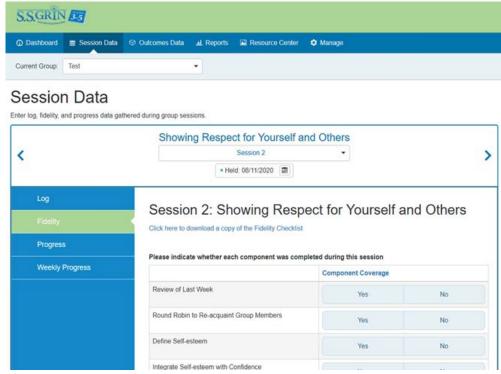

Figure 1: Sample Screenshot of IMPACT Data Entry Screen for S.S.GRIN

Table 1. Intercorrelations for IMPACT Ratings at POST (Note. Bolded statistics indicate correlations among areas within an implementation outcome area. *$p < .01$, **$p < .001$, ***$p < .0001$)

Area Rated	Sat	Val	Sup	Fit	Comp	Cap	Fea	IQ	Usa	Cont	C2C
Acceptability											
Satisfaction	--										
Value	**.77*****	--									
Superior Innovation	**.82*****	**.70*****	--								
Appropriateness											
Setting fit	.82***	.71***	.74***	--							
Work Compatability	.67***	.73***	.80***	**.65*****	--						
Implementation Support Capacity	.53***	.54***	.51***	**.55*****	**.37***	--					
Feasibility											
Feasibility	.77***	.75***	.71***	.80***	.72***	.51***	--				
Interface quality	.65***	.63***	.61***	.57***	.58***	.45**	**.70*****	--			
System usability	.50***	.47***	.44***	.45***	.41**	.41**	**.67*****	**.66*****	--		
Adoption											
Continue use	.70***	.70***	.72***	.74***	.66***	.43***	.68***	.44***	.34*	--	
Commitment to change	.62***	.60***	.73***	.64***	.63***	.37*	.64***	.53***	.33*	**.80*****	--
Recommend use	.67***	.76***	.66***	.57***	.64***	.52***	.68***	.68***	.37***	**.60*****	**.66*****

of the study and whether usage influenced ratings of implementation outcomes at the end of the study period. We expected greater use of IMPACT to result in more positive views of the platform. Lastly, we explored whether results differed significantly by demographic characteristics of our providers (gender, race, ethnicity). We did not expect differential patterns of results for any demographic subgroup.

Methods

Procedures: This study took place over the 2019-2020 school year. In the fall, we worked with several school districts that routinely use S.S.GRIN as a social-emotional learning (SEL) intervention for their students. Following district-level approval, districts shared information about the IMPACT evaluation study with S.S.GRIN providers. To participate, providers agreed to (a) deliver at least one S.S.GRIN group over the course of the school year and (b) use IMPACT to enter and track data for their S.S.GRIN group(s). Once eligibility was confirmed, active consent for participating providers was obtained.

Prior to delivering S.S.GRIN with students, providers completed in-person and/or online training in delivery of the S.S.GRIN evidence-based intervention, as required by their district. On average, providers completed a total of six hours of preimplementation training. Providers were also required to pass the online certification test for S.S.GRIN. On average, providers achieved 93% proficiency (range 80% to 100%). In addition, providers who volunteered to participate in the IMPACT evaluation research completed a one-hour training workshop (in-person or via webinar) in the use of IMPACT specifically.

Following training and before delivery of S.S.GRIN (PRE time point), providers completed an online survey to provide demographic and background information. From that point forward, providers selected students and implemented S.S.GRIN groups as they typically would and used IMPACT for data entry and analysis of session fidelity, student progress, and student outcomes. Once providers indicated their implementation of S.S.GRIN was complete, and they did not expect to run any additional groups with students for the remainder of the school year, providers completed POST online survey measures.

Participants

Attrition analyses: Ninety-nine providers completed the PRE assessment survey. Of those, 21 failed to complete POST data collection. Attrition was largely a result of school closures due to COVID-19, cited by 10 providers (48%) as the reason for leaving the study, and another

Table 2. Means (standard deviations) for IMPACT Ratings at POST ([a]7-point scale from 1 (strongly disagree) to 7 (strongly agree), [b]5-point scale from 1 (Poor) to 5 (excellent), [c]5-point scale from 1 (none) to 5 (very high), [d]4-point scale from 1 (not at all easy) to 4 (very easy), [e]4-point scale from 1 (not at all likely) to 4 (very likely) *$p < .0001$

Area Rated	Mean (SD)	$T_{(77)}$ Test Statistic
Satisfaction[a]	5.49 (1.01)	13.07*
Value[b]	3.75 (.99)	6.68*
Superior Innovation[a]	5.36 (.95)	12.59*
Setting fit[a]	5.43 (1.07)	11.72*
Work Compatibility[a]	5.09 (.99)	9.74*
Implementation Support Capacity[c]	3.48 (.72)	5.85*
Feasibility[a]	5.47 (1.10)	11.73*
Interface quality[a]	5.58 (1.07)	13.12*
System usability[d]	3.00 (.70)	12.08*
Continue use[a]	5.32 (1.36)	8.53*
Commitment to change[a]	5.54 (1.13)	12.05*
Recommend use[e]	3.06 (.92)	10.36*

five (24%) dropped out for personal reasons (e.g., left school system, maternity leave). To investigate whether selective attrition occurred, we conducted Chi-square analyses to compare demographic characteristics (i.e., district, gender, race, ethnicity, age, education, position, years of service) of those who did versus did not attrit. Using a Bonferroni-corrected *P*-value of .01, no significant differences were found. Thus, results did not indicate selective attrition for different demographic subgroups in our study.

Sample characteristics: The longitudinal data set was composed of 78 providers who worked across 49 schools in four districts serving primarily urban and suburban communities. The majority of providers were female (91%). The racial distribution was 66.7% White, 16.7% Black, 7.7% American Indian/Alaska Native, 6.4% Multiracial, and 2.5% Asian American, with 19% of providers reporting Hispanic/Latinx heritage. Providers represented a wide age range (18 to 66 years), with a median age of 32 to 38 years. Forty percent identified as a social worker, 32% as a counselor, 22% as a psychologist, and 6% as a teacher or other school staff.

Measures

Implementation data experience: At PRE, providers indicated whether they had any prior experience collecting or tracking session fidelity, student progress, and/or student outcomes data for S.S.GRIN or any other SEL program. Those providers with prior implementation

Table 3. Univariate Statistics and Standardized Beta Weights for Dislike of Implementation Data Tracking at PRE Predicting POST Ratings by Prior Experience Type. (*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$)

Prior Experience Type			
Area Rated	$F_{(1,57)}$ Test Statistic	Other SEL Std β	SSGRIN Std β
Satisfaction	8.12**	.01	.64****
Value	5.23*	-.04	.52***
Superior Innovation	10.86**	-.10	.65****
Setting fit	5.51*	.06	.58***
Work Compatibility	5.13*	-.03	.49**
Implementation Support Capacity	ns	.05	.38*
Feasibility	ns	.15	.39*
Interface quality	5.40*	-.13	.46**
System usability	ns	-.01	.43*
Continue use	5.42*	.05	.67****
Commitment to change	5.86*	-.21	.40*
Recommend use	6.48*	-.12	.49**

data experience then completed follow-up questions regarding (a) the specific methods used for these purposes in the past and (b) the degree to which they disliked entering and tracking fidelity, progress, and outcomes data (three items). Dislike ratings were made on a 7-point scale from 1= 'Strongly Disagree' to 7= 'Strongly Agree', and a mean score was created across items.

Implementation outcomes: At POST, providers rated their experience with and perceptions of the IMPACT platform for S.S.GRIN. Four implementation outcome areas were assessed using three subscales per area. Subscales were selected in an effort to assess distinct aspects of each outcome area and thereby provide a thorough investigation of each implementation outcome for IMPACT. For each subscale, a mean score was computed across respective items.

Acceptability: The degree to which providers found use of IMPACT agreeable, palatable, and satisfactory was assessed via Satisfaction, Superior Innovation, and Value subscales. *Satisfaction* items (7) evaluated the degree to which providers were satisfied with specific aspects of working with the IMPACT platform, such as "I am satisfied with the process of entering fidelity data using IMPACT." *Superior Innovation* items (5) assessed the degree to which providers viewed IMPACT as superior to other methods that could be used for the same purposes (e.g., "IMPACT is a superior way to track fidelity for S.S.GRIN"). Satisfaction and Superior Innovation items were rated on a 7-point scale from 1 (Strongly Disagree) to 7 (Strongly Agree). *Value* items (9) assessed providers' perceptions of IMPACT's overall value for helping them achieve high-quality implementation for S.S.GRIN (e.g., "For helping you identify gaps in your implementation fidelity") and were rated on a 5-point scale from 1 (None) to 5 (Excellent). Cronbach's alpha indicated excellent internal consistency across respective items for each subscale (Satisfaction $\alpha = .90$, Superior Innovation $\alpha = .85$, Value $\alpha = .96$).

Appropriateness: The degree to which providers found use of IMPACT relevant, compatible, and a good fit for the work they do was assessed via Setting Fit, Work Compatibility, and Implementation Support Capacity subscales. *Setting Fit* items (8) evaluated the degree to which providers believed use of IMPACT fit within their service delivery setting, such as "Using IMPACT's fidelity tracking tools is practical to do in my current setting Transpose:." *Work Compatibility* items (5) assessed the degree to which use of IMPACT for S.S.GRIN is compatible with other systems used by providers in their work (e.g., "Useful addition to other existing systems at my practice setting" and "Complements data available via other LMS/EH systems I use"). Fit

and Compatibility items were rated on a 7-point scale from 1 (Strongly Disagree) to 7 (Strongly Agree). *Implementation Support Capacity* items (3) assessed providers' perceived skill level for accomplishing fidelity, progress monitoring, and outcomes tracking via IMPACT at the end of the study period (e.g., "my skill level for tracking implementation fidelity for S.S.GRIN") and were rated on a 5-point scale from 1 (None) to 5 (Very High). Cronbach's alpha indicated good internal consistency for each subscale (Setting Fit $\alpha = .92$, Work Compatibility $\alpha = .84$, Implementation Support Capacity $\alpha = .76$).

Feasibility: The degree to which providers could successfully use IMPACT for S.S.GRIN was assessed via Feasibility, Interface Quality, and Ease of Use subscales. The *Post Study System Usability Questionnaire* (PSSUQ; Lewis, 2002) was used to rate IMPACT's overall feasibility and interface quality. *Feasibility* items (10) evaluated providers' perceptions of IMPACT's overall feasibility of use, such as "IMPACT is a time-efficient system". *Interface Quality* items (4) assessed provider's perceptions of the overall quality of IMPACT's software user interface (UI), such as "IMPACT has all the functions and capabilities I would expect it to have". Feasibility and Interface Quality items were rated on a 7-point scale from 1 (Strongly Disagree) to 7 (Strongly Agree). *Ease of Use* items (7) assessed providers' overall perceptions of how easy it was for them to use the software system, such as "I knew where to find the information I needed or wanted" and were rated on a 4-point scale from 1 (Not at all Easy) to 4 (Very Easy). Cronbach's alpha indicated high internal consistency for each subscale (Feasibility $\alpha = .94$, Interface Quality $\alpha = .89$, Ease of Use $\alpha = .95$).

Adoption: Providers' intent to use IMPACT to support their S.S.GRIN implementation in the future was assessed via Continue Use, Commitment to Change, and Recommend Use subscales. *Continue Use* items (5) evaluated providers' level of intent to continue using IMPACT for its designed purposes for S.S.GRIN, such as "I intend to use IMPACT to track outcomes for students in S.S.GRIN after this study Transpose: ." The Affective subscale (seven items) of the *Commitment to Change* measure (Herscovitch & Meyer, 2002) was used to assess the degree to which providers believed their organization, and those delivering S.S.GRIN in their organization, should move to using IMPACT, such as "I believe in the value of this change." Items for these two subscales were rated on a 7-point scale from 1 (Strongly Disagree) to 7 (Strongly Agree) and Cronbach's alpha indicated excellent internal consistency (Continued use $\alpha = .95$, Commitment to Change $\alpha = .93$). One item from the PSSUQ (Lewis, 2002) was used to assess the degree to which providers would recommend use of IMPACT to others delivering S.S.GRIN ("How likely are you to recommend use of IMPACT for S.S.GRIN to a colleague or peer?"). This item was rated on a 4-point scale from 1 (Not at all likely) to 4 (Very likely).

IMPACT usage: Over the course of the study, usage metrics were collected by the software, including the dates of first and last logins to the system, total number of logins, and total number of minutes spent logged into the system. The number and type of technical assistance requests were also tracked.

Results

Preliminary analyses: Table 1 presents the intercorrelations among the 12 implementation outcome subscales included in this study. All subscales were significantly related to one another, but no subscale appeared to be redundant with another (i.e., no correlation greater than .90), supporting each as a distinct, but related aspect of implementation. As expected, the three subscales within each area were significantly interrelated and showed high internal consistency (Cronbach's alphas ranging from .76 for Appropriateness to .89 for Acceptability). However,

there were also interesting cross-area correlations. For example, while the Feasibility subscale was highly correlated the other two subscales under the Feasibility outcome area ($r = .67$ with System Usability and $r = .70$ with Interface Quality), its highest bivariate correlation was with Setting Fit ($r = .80$) under the Appropriateness outcome area.

The highest bivariate correlations involved the Satisfaction subscale ($r = .82$ with Superior Innovation and Setting Fit subscales). Thus, provider satisfaction with the innovation was highly correlated with seeing IMPACT as superior to other methods to accomplish the same tasks and seeing use of IMPACT as fitting into the service delivery setting. Interestingly, the lowest bivariate correlations were between the software Ease of Use subscale and the three Adoption subscales ($r = .33$ with Commitment to Change, $r = .34$ with Continue Use, and $r = .37$ with Recommend Use). Therefore, while being seen as easy to use was related to each adoption subscale, software usability was not as highly related to providers' likelihood of adoption or continued use beyond this research study.

Mean POST ratings: Table 2 displays the means and standard deviations for provider ratings of IMPACT at the end of the study period (POST). Given varying scales were used, the table includes reference to the specific scale used for each area assessed. In order to test whether these reported mean subscale scores were significantly higher than the mid-point for its respective scale (e.g., higher than 3 for a 5-point scale), single sample t test were conducted. All t-test statistics were significant at the .0001 p-level, indicating that each area was rated significantly higher than the neutral/average midpoint value for that scale. While all areas were significant, the t-test statistics were particularly high for the Satisfaction, Superior Innovation, and Interface Quality subscales.

Prior implementation data experience: Sixty-eight providers (87%) reported experience collecting, entering, and/or tracking implementation data (fidelity, progress, outcomes) prior to the study, and 10 providers (13%) reported no prior experience. Of the 68 experienced providers, 31 had implementation data experience specifically for S.S.GRIN, and 37 reported prior experience for a different SEL program. Multivariate analysis of variance (MANOVA) results for the prediction of POST ratings indicated no significant difference in ratings by type of prior implementation data experience.

Prior methods used: The 68 providers who reported prior experience collecting and tracking implementation data were asked to list the method(s) used to accomplish these tasks. The most common method was paper-and-pencil, with 79% reporting having used this method for one or more tasks. Thirty percent reported using Excel to assist in data collection, scoring, and/or tracking, and 34% reported using some other software, such as an outcome measure scoring program. To examine whether POST ratings differed by prior methods (paper, Excel, other software), separate MANOVAs were conducted using that method (Y/N) predicting POST ratings. No significant multivariate main effect was present for any method.

Dislike of implementation data tracking: At PRE, providers with prior experience were asked to rate their overall dislike for engaging in implementation data collection and tracking activities ($M = 4.40$, $SD = 1.54$). An analysis of variance (ANOVA) revealed disliking did not differ significantly by type of prior method used (paper, Excel, other software). In the prediction of POST ratings of IMPACT, there was no significant multivariate main effect for dislike at PRE. However, there was a significant multivariate interaction effect for dislike by implementation data experience type (for S.S.GRIN vs for another SEL program; $F_{(12,46)} = 2.05$, $p < .05$). Post hoc regression analyses were run to determine the standardized beta weights for the prediction of dislike

on POST ratings separately for each data experience type. As Table 3 summarizes, for those providers with prior experience collecting and tracking implementation data specifically for S.S.GRIN, the more they disliked these activities, the higher their POST ratings of all aspects of IMPACT. No significant differences were present for providers with prior implementation data experience for some other SEL program.

IMPACT usage: On average, providers used IMPACT for 18.57 weeks ($SD = 5.86$ weeks). The total number of logins over the study period averaged 23.12 ($SD = 17.11$), and, on average, providers spent a total of 96.58 minutes logged on to the system ($SD = 562.78$). The number of logins and number of minutes were not significantly correlated with one another, which is not surprising given the variety of ways in which users can use such a system and that minutes are calculated for total time logged into the system, regardless of whether the user was actively using the system or otherwise engaged (e.g., stepped away, on the telephone). Overall, providers made very few requests by email or phone for technical assistance (TA) with the software. The range of TA requests by providers was 0 to 4 over the study period, with an average of fewer than one per provider ($M = 0.90$, $SD = 1.01$). The most common TA request was for help logging in (e.g., forgot password).

The correlation between dislike of implementation data tracking at PRE was significantly negatively related to the number of logins ($r = -.34$, $p < .01$), but not the number of minutes. Thus, the more providers disliked implementation data tasks at the start of the study, the fewer times they logged into IMPACT over the course of the study. The MANOVA predicting POST ratings for IMPACT usage showed a significant multivariate main effect for number of logins ($F_{(12,64)} = 2.83$, $p < .01$). Post-hoc univariate analyses indicated Implementation Support Capacity ($F_{(1,75)} = 9.18$, $p < .01$, Std $\beta = .32$) and Recommend Use ($F_{(1,75)} = 5.55$, $p < .05$, Std $\beta = .26$) were significantly higher at POST for those providers who logged into IMPACT more frequently. No significant effect on POST ratings was found for number of minutes.

Demographic sub-group analyses: Follow-up analyses were conducted to explore whether the pattern of findings may differ by provider demographic characteristic. For each area assessed as described above, parallel analyses were run separately to test for differences by gender (male or female), racial (Black, multiracial, or other) subgroups, and ethnic (Hispanic/Latinx or not) subgroups. Chi-square analyses were conducted for categorical variables, and main and interaction effects for demographic characteristic were included in analyses of variance predicting POST ratings and usage metrics.

Across all sets of analyses, no significant differences were found for racial or ethnic subgroups, and only one significant effect involved gender. Specifically, Chi-square analyses for prior methods used for tracking implementation data (i.e., paper-and-pencil, Excel, or other software) showed that female providers were more likely to have used paper-and-pencil methods compared to male providers (83% vs 33%; $\chi^2_{(1)} = 8.55$, $p < .01$). Otherwise, the patterns of findings across analyses did not differ by any demographic characteristic of participating providers.

Discussion

As expected, results of this evaluation provided support for the utility and value of the IMPACT implementation support platform. Following use of IMPACT for entry and tracking of implementation data for the S.S.GRIN evidence-based program, providers rated the software system positively for all implementation outcome areas. Specifically, providers reported that IMPACT was valuable for helping them achieve high-quality implementation of S.S.GRIN, and that they

were satisfied with IMPACT for accomplishing implementation data tasks, seeing IMPACT as superior to other methods that could be used for the same purposes (i.e., high ratings of acceptability).

Providers also rated IMPACT as highly relevant to and compatible with the work they do as providers of SEL services in schools and saw IMPACT as fitting well within their service delivery setting (i.e., high ratings of appropriateness). Providers further rated IMPACT's software user interface as of high quality and easy and feasible to use (i.e., high ratings of feasibility). And lastly, providers expressed a commitment to continue using IMPACT to support their implementation of S.S.GRIN after the end of the study, reported they would be likely to recommend use of IMPACT to others, and believed those delivering S.S.GRIN in their organization should move to using IMPACT (i.e., high ratings for adoption). While all areas were rated significantly above average, IMPACT was rated particularly highly in terms of provider satisfaction, superiority over alternative methods, and software usability.

We expected providers' prior experiences collecting and tracking implementation data would not influence their ratings of IMPACT. In general, findings provided support for this hypothesis. When comparing ratings for those with and without prior implementation data experience, there were no group differences in ratings of IMPACT. Similarly, for those with experience, the specific prior method used did not influence ratings of IMPACT. Amazingly in this day and age, the most common method reported was paper and pencil (79%). Furthermore, when we looked at the degree to which providers reported disliking collecting and tracking implementation data before engaging with IMPACT, this PRE disliking did not influence POST ratings of IMPACT, except for those providers who had previous implementation data experience specifically for S.S.GRIN.

Therefore, in general, all sets of providers rated IMPACT positively in all areas at POST. However, for those providers who had used some other method(s) for collecting and tracking fidelity, progress, and/or outcomes data specifically for S.S.GRIN *and* they disliked those activities, POST ratings of IMPACT were particularly likely to be positive. While this interaction effect was significant for all areas, standardized beta weights showed the effect was greatest for seeing IMPACT as a superior alternative, expressing greater satisfaction with IMPACT, and indicating higher likelihood of continued use. Given this evaluation was of an IMPACT application specifically for S.S.GRIN, and POST ratings were specific to using IMPACT for S.S.GRIN, it makes sense that ratings by providers with prior experience with implementation data for S.S.GRIN might be most impacted. These providers may be most likely to recognize differences offered by a new system and, if judged as improvements, be most likely to value the new system.

Along these same lines, negative perceptions of prior implementation data methods were also found to influence the actual usage of IMPACT. Specifically, the more a provider expressed dislike (at PRE) for collecting and tracking fidelity, progress, and/or outcomes data, the fewer times they logged into the IMPACT software system over the study period. This significant negative association held across providers with any prior implementation data experience regardless of whether that experience was for S.S.GRIN or some other SEL program. This finding suggests negative biases may undermine introduction of an innovation. However, if users engage with an innovation despite any pre-existing biases, and if the innovation is seen as superior, then attitudes towards the new innovation can be particularly positive, and thereby actually support adoption of the innovation.

We expected greater IMPACT usage over the course of this study to result in more positive ratings at POST. We found support for this

hypothesis in two implementation outcome areas. First, the more times a provider logged in to IMPACT, the higher they rated their skills and abilities for collecting and tracking implementation data (fidelity, progress, and outcomes) at POST. In other words, providers saw their implementation support capacity as significantly greater when they used IMPACT to a larger extent. Second, providers reported a higher likelihood of recommending IMPACT to a peer or colleague when they had used the system more during the study period.

This study benefited from a varied sample of providers, which enabled us to investigate whether the patterns of results differed for specific subgroups. As expected, no significant sub-group differences were found for any demographic characteristic (i.e., gender, race, ethnicity) in the prediction of POST ratings of implementation outcomes, and IMPACT usage was similar across all demographic sub-groups. Therefore, use of the IMPACT implementation support system was found to be similarly beneficial for the full range of participating providers.

Limitations

Unfortunately, due to the COVID-19 pandemic, our evaluation study was cut short by several months. A number of our providers could not complete their S.SGRIN group(s) with students because of school closures and, therefore, could not complete data collection at POST. However, despite a higher than expected attrition rate (21%), no evidence of selective attrition was found. Therefore, the loss of provider participants did not appear to be attributable to demographic or background characteristics.

This evaluation focused on a customized application of IMPACT tailored to the S.S.GRIN EBPP. The IMPACT software system maintains a similar UI, set of features and functions, and database management structure for any EBPP built on the platform. Customization is accomplished by tailoring the look and feel (e.g., colors), data collection specifics (e.g., measures, schedule, respondent types), and implementation support resources (e.g., training videos) for a specific EBPP. We would expect results of this evaluation to generalize to any other EBPP supported by IMPACT. It is possible, however, that evaluation results could be different for an EBPP delivered in a different format (such as one-on-one rather than group-based) or focused on different clinical outcomes (such as mental health rather than SEL). It will be important to replicate the findings of this evaluation study through future research with different types of EBPPs in order to ensure results are indeed generalizable.

In addition, this evaluation study is limited methodologically by its focus on ratings only at the POST time point. While this study provides initial evidence of positive evaluations after experience with the system, looking at change over time as a function of that experience would strengthen conclusions about the implementation outcomes for IMPACT. Further, this study was only able to compare naturally occurring groups of providers (i.e., those with and without prior implementation data experience), but it was not feasible for our participating school districts to randomly assign providers to different implementation support systems. Longitudinal and randomized control group studies would greatly strengthen the conclusions that can be drawn about IMPACT compared to alternative implementation data collection and tracking methods.

Another methodological limitation of this study was the fairly gross measurement of system usage. The number of logins and number of minutes logged in were captured, but we did not capture information about user behaviors while logged in. Users can do a wide array of activities through the IMPACT system and may be logged in but not

actively engaged in any activity (e.g., on the phone, stepped away from the computer). A more granular examination of user behavior, such as how much time was spent in a particular activity or the number of times a particular task was completed, would be important to move this field further in understanding how engagement in software impacts implementation outcomes.

Future directions

This study purposefully included a variety of subscales intended to operationalize various aspects of the acceptability, appropriateness, feasibility, and likelihood of adoption outcomes based on the Proctor et al. (2011) implementations outcome framework. The interrelations among the 12 subscales included in this study provided support for this framework and indicated some interesting cross-outcome associations that could inform our understanding of how specific aspects of an implementation system influence one another. For example, perceptions of feasibility and appropriateness may be more closely tied than previously thought, and it may be useful to consider software usability as a separate construct within the implementation outcomes framework. In addition, it may be useful to consider how implementation outcomes influence one another. For example, in our study, seeing IMPACT as a superior innovation was strongly related to providers' commitment to using IMPACT in the future. Perhaps noting clear advantages over existing methods encourages adoption of an innovation. The relative importance of specific implementation outcome areas is not well understood currently. Further study is needed to fully understand each construct and the interplay among the various implementation outcome areas over time.

Conclusion

This study suggested negative provider preconceptions about collecting and tracking implementation data may present a challenge for building organizational capacity to scale EBPPs. In our study, negative perceptions were associated with lower usage of IMPACT, suggesting lower engagement and motivation to use the innovation when providers entered the study with negative pre-existing perceptions of implementation data tracking. However, providers who entered the study with such preconceptions for the target EBPP (S.S.GRIN) showed the most positive evaluations after using IMPACT, suggesting perhaps that a good experience with an innovation can overcome negative preconceptions. Future work that directly addresses providers' past experiences with and reactions to prior systems could shed light on how these pre-conceptions may influence adoption and use of innovations for implementation data tracking.

Funding

Conflicts of interest

Lead author Dr. DeRosier has an intellectual property interest and copyright interest in IMPACT. She is the President and Chief Executive Officer of 3C Institute, which developed IMPACT. She stands to gain monetarily if IMPACT is marketed and sold.

Acknowledgements

References

1. Aarons GA, Sklar M, Mustanski B, Benbow N, Brown CH, et al. (2017) "Scaling-out" evidence-based interventions to new populations or new health care delivery systems. *Implementation Science* 12: 111. [Crossref]

2. Berkel C, Gallo CG, Sandler IN, Mauricio AM, Smith JD, et al. (2019) Redesigning implementation measurement for monitoring and quality improvement in community delivery settings. *Journal of Primary Prevention* 40: 111-127. [Crossref]

3. Bickman L (2008) A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry* 47: 1114-1119. [Crossref]

4. Bruns EJ, Suter JC, Force MM, Burchard JD (2005) Adherence to wraparound principles and association with outcomes. *Journal of Child and Family Studies* 14: 521-534.

5. Bruns EJ, Suter JC (2010) Summary of the wraparound evidence base. The resource guide to wraparound. Portland, OR: National Wraparound Initiative.

6. DeRosier ME (2004) Building relationships and combating bullying: Effectiveness of a school-based social skills group intervention. *Journal of Clinical Child and Adolescent Psychology* 33: 125-130. [Crossref]

7. DeRosier ME (2007) Peer victimized and rejected children: Promoting school-based adjustment through social skills intervention. In J.E. Zins, M.J. Elias, & C.A. Maher (Eds.), Handbook of prevention and intervention in peer harassment, victimization, and bullying (pp. 257-271). New York: Haworth Press.

8. DeRosier ME (2007) Social Skills GRoup INtervention (S.S.GRIN): Professional Manual (4th ed.). Cary, NC: 3-C Institute for Social Development.

9. DeRosier ME (2019) Three critical elements for real-time monitoring of implementation and adaptation of prevention programs. *Journal of Primary Prevention* 40: 129-135. [Crossref]

10. DeRosier ME, Marcus SR (2005) Building friendships and combating bullying: Effectiveness of S.S.GRIN at one-year follow-up. *Journal of Clinical Child and Adolescent Psychology* 34: 140-150.

11. Durlak JA, DuPre EP (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. [Crossref]

12. Effland VS, Walton BA, McIntyre JS (2011) Connecting the dots: Stages of implementation, wraparound fidelity and youth outcomes. *Journal of Child and Family Studies* 20: 736-746.

13. Fagan AA, Bumbarger BK, Barth RP, Bradshaw CP, Cooper BR, et al. (2019) Scaling up evidence-based interventions in US Public Systems to prevent behavioral health problems: Challenges and opportunities. *Prevention Science* 20: 1147-1168. [Crossref]

14. Fixsen D, Blasé K, Metz A, Van Dyke M (2013) Statewide implementation of evidence-based programs. *Exceptional Children* 79: 213-230.

15. Gilbody SM, House AO, Sheldon TA (2002) Psychiatrists in the UK do not use outcomes measures: National survey. *British Journal of Psychiatry* 180: 101-103. [Crossref]

16. Hatfield DR, Ogles BM (2004) The use of outcome measures by psychologists in clinical practice. *Professional Psychology Research and Practice* 35: 485-491.

17. Herscovitch L, Meyer JP (2002) Commitment to organizational change: Extension of a three-component model. *Journal of Applied Psychology* 87: 474-487. [Crossref]

18. Jensen-Doss A, Weisz JR (2008) Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology* 76: 711-722. [Crossref]

19. Kazak AE, Hogwood K, Weisz JR, Hood K, Kratochwill TR, et al. (2010) A meta-systems approach to evidence-based practice for children and adolescents. *American Psychologist* 65: 85-97. [Crossref]

20. Kazdin AE (2008) Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist* 63: 146-159. [Crossref]

21. Lambert MJ, Whipple JL, Smart DW, Vermeersch DA, Nielsen SL, et al. (2001) The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research* 11: 49-68.

22. Lewis JR (2002) Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction* 14: 463-488.

23. McHugh RK, Barlow DH (2010) The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist* 65: 73-84. [Crossref]

24. McMillen J, North K, DeRosier ME, Davis N (2010) Social Skills GRoup INtervention for Early Childhood (S.S.GRIN-EC): Professional Manual (2nd ed.). Cary, NC: 3-C Institute for Social Development.

25. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, et al. (2011) Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research* 38: 65-76. [Crossref]

26. Weaver C, DeRosier ME (2019) Commentary on Scaling-Up Evidence-Based Interventions in Public Systems. *Prevention Science* 20: 1178-1188. [Crossref]

27. Weisz JR, Donenberg GR, Han SS, Weiss B (1995) Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology* 63: 688-701. [Crossref]

28. Weisz JR, Hawley KM, Doss AJ (2004) Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child Adolescent Psychiatric Clinics of North America* 13: 729-815. [Crossref]

29. Weisz JR, Sandler IN, Durlak JA, Anton BS (2005) Promoting and protecting youth mental health through evidence-based prevention and treatment. *American Psychologist* 60: 628-48. [Crossref]