Developing a fidelity measure of early intervention programs for children with neuromotor disorders

MIHEE AN[1] | JAYDEN NORD[2] | NATALIE A KOZIOL[2] | STACEY C DUSING[3] | AUDREY E KANE[4] | MICHELE A LOBO[5] | SARAH W MCCOY[6] | REGINA T HARBOURNE[7]

1 Department of Physical Therapy, Kaya University, Gimhae, Republic of Korea. **2** Nebraska Center for Research on Children, Youth, Families and Schools, University of Nebraska, Lincoln, NE; **3** Division of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, CA; **4** Department of Physical Therapy, Virginia Commonwealth University, Richmond, VA; **5** Department of Physical Therapy, University of Delaware, Newark, DE; **6** Department of Physical Therapy, University of Washington, Seattle, WA; **7** Department of Physical Therapy, Duquesne University, Pittsburgh, PA, USA.

Correspondence to Regina T Harbourne, Department of Physical Therapy, 600 Forbes Avenue, Duquesne University, Pittsburgh PA, USA. E-mail: harbourner@duq.edu

PUBLICATION DATA
Accepted for publication 31st August 2020.
Published online 13th October 2020

ABBREVIATIONS

START-Play	Sitting Together and Reaching to Play
uEI	Usual early intervention

AIM To describe the development of an intervention-specific fidelity measure and its utilization and to determine whether the newly developed Sitting Together and Reaching to Play (START-Play) intervention was implemented as intended. Also, to quantify differences between START-Play and usual early intervention (uEI) services.

METHOD A fidelity measure for the START-Play intervention was developed for children with neuromotor disorders by: (1) identifying key intervention components, (2) establishing a measurement coding system, and (3) testing the reliability of instrument scores. After establishing acceptable interrater reliability, 103 intervention videos from the START-Play randomized controlled trial were coded and compared between the START-Play and uEI groups to measure five dimensions of START-Play fidelity, including adherence, dosage, quality of intervention, participant responsiveness, and program differentiation.

RESULTS Fifteen fidelity variables out of 17 had good to excellent interrater reliability evidence with intraclass correlation coefficients (ICCs) ranging from 0.77 to 0.95. The START-Play therapists met the criteria for acceptable fidelity of the intervention (rates of START-Play key component use ≥0.8; quality ratings ≥3 [on a scale of 1–4]). The START-Play and uEI groups differed significantly in rates of START-Play key component use and quality ratings.

INTERPRETATION The START-Play fidelity measure successfully quantified key components of the START-Play intervention, serving to differentiate START-Play from uEI.

Intervention fidelity refers to faithful and accurate implementation of an intended intervention. It is a multidimensional construct with five dimensions: adherence, dosage, quality of intervention delivery, participant responsiveness, and program differentiation (Fig. 1).[1–5] Adherence measures the accurate delivery of the key components of an intervention as it was designed. Dosage refers to the amount of a specific intervention delivered, measured in terms of time, frequency, or rate. Quality of intervention delivery concerns the manner in which therapists implement the intervention using the overall processes or strategies designed by the developers. Participant responsiveness involves the extent to which participants respond to, or are engaged in, the intervention. Program differentiation indicates whether the characteristics of a target intervention can be differentiated from those of other interventions. Program differentiation can include items comparing groups from the other four dimensions of fidelity; differentiation between approaches is important to identify the active elements essential for a successful intervention. Adequate measurement and reporting of the multiple dimensions of intervention fidelity ensures the accurate presentation, implementation, and examination of the target intervention, and improves the replicability of the intervention.[1,2] However, studies using fidelity measurement in rehabilitation generally focus on only adherence and not the other dimensions of fidelity measurement.[6–10]

The Sitting Together and Reaching to Play (START-Play) intervention aims to advance motor and cognitive skills in infants with neuromotor disorders.[11,12] A consortium of pediatric physical therapists, psychologists, education researchers, and interventionists developed the START-Play intervention based on effectiveness research and child development theories supporting grounded cognition.[12] The START-Play intervention emphasizes the importance of developing early motor skills (sitting and reaching) alongside the progression of problem-solving abilities, thus advancing learning and overall development (see Appendix S1, online supporting information, for comparison of START-Play and usual early intervention [uEI]). Therefore, the START-Play intervention focuses on promoting therapeutic behaviors linking infants' sitting,

reaching, and problem-solving skills, and building the capacity of parents to continue the linkage of these areas.

When a new intervention is compared to existing practice, it is important to use a fidelity measure that reflects the unique features or key components of the new intervention. A fidelity measure for the START-Play intervention was developed as part of a randomized controlled trial comparing infants who received uEI services (uEI group) to infants who received the START-Play intervention in addition to uEI (START-Play group).[11] UEIs are federally funded services received by infants participating in the START-Play randomized controlled trial and may vary in type and amount. UEI was not defined or explained other than in the context of our measure, with early intervention provided to children because of motor delays, specifically delays in sitting and reaching. There was no intention to include or exclude a specific model or approach to early intervention and an eclectic approach appeared to predominate. No defined infant rehabilitation approach for early intervention was noted for any child in this study. The aim and intent of this fidelity measure was to examine if the START-Play therapists were implementing the START-Play intervention as intended (adherence) and if the START-Play intervention was actually different from the child's uEI (program differentiation). For program differentiation, the content and amount of intervention (adherence and dosage) and intervention process (quality of intervention delivery and participants responsiveness) were compared between the two groups. This paper highlights the two-phase process conducted to develop the intervention-specific fidelity measure and the implementation of the measure to quantify adherence and program differentiation during the randomized controlled trial.

What this paper adds

- The Sitting Together and Reaching to Play (START-Play) fidelity measure is the first intervention-specific measure used in pediatric rehabilitation.
- The measure is reliable between raters in evaluating behaviors relevant to START-Play intervention.
- The measure differentiates the START-Play intervention from usual early intervention by quantifying differences in five dimensions of fidelity.

METHOD

Phase I: development and reliability of the fidelity measure

Seventy-nine therapists from four different regions of the United States (Northwest, Midwest, Northeast, Southeast) participated in the START-Play clinical trial. Ethical approval was obtained from a central (Duquesne University) or specific (Virginia Commonwealth University) institutional review board. Informed written consent for videotaping intervention sessions was obtained from parents and therapists in both groups. The therapists in the

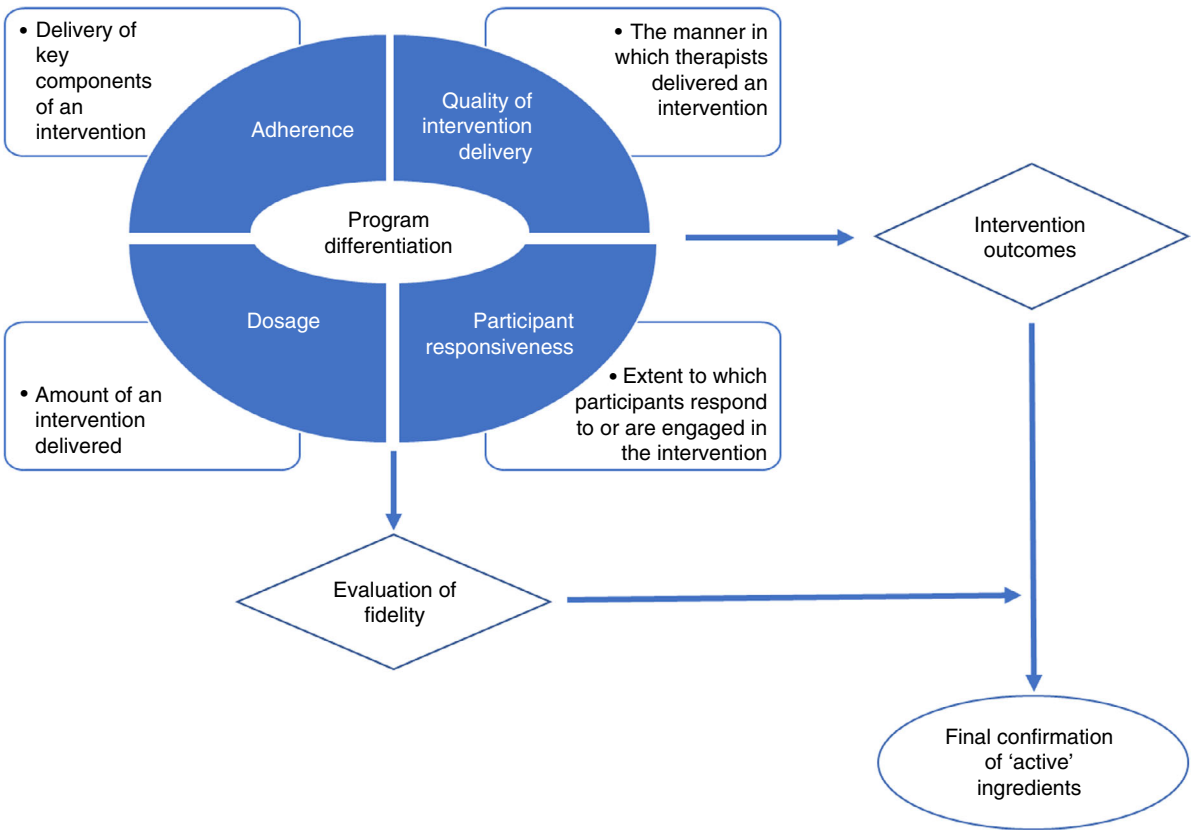

Figure 1: Multidimensional construct of the Sitting Together and Reaching to Play (START-Play) intervention fidelity. [Colour figure can be viewed at wileyonlinelibrary.com]

START-Play group (n=14) received 3 days of training by the principal investigator of the START-Play research team and had ongoing oversight by an on-site principal investigator. The therapists were instructed to provide the START-Play intervention consistently in collaboration with a parent or caregiver and had access to their site principal investigator for feedback and advice. The other group of therapists were early intervention (Individuals with Disabilities Education Act Funded) providers for infants randomly assigned to the control (uEI) group. If a parent of an infant who was assigned to the uEI group agreed, the infant's usual care therapist was invited to participate. Forty-nine therapists (14 in START-Play, 35 in uEI) consented to videotaping and were used in the fidelity analysis. Demographic information for all therapists who consented to video analysis, including age, sex, level of education, professional title, years of experience, and ethnicity is presented in Appendix S2 (online supporting information).

The development of the START-Play fidelity measure was a three-step process: (1) identifying key components, (2) establishing a measurement system (i.e. how to measure the key components and how to determine if the intervention was implemented with acceptable fidelity), and (3) evaluating reliability evidence.[13] The process of developing an intervention-specific, multi-dimensional fidelity measure is detailed in a perspective paper and briefly described below.[6]

Step 1: identifying key components

Three members of the START-Play research team (RH, SC, SS) participated in steps 1 and 2. Based on the theory of change supporting the START-Play intervention and detailed analysis of sample videos of intervention sessions representing START-Play and uEI, key components were identified and operationally defined (Table S1, online supporting information).[11,14] The key components were intended to examine the extent to which therapists deliver the intervention (six START-Play behaviors) and avoid activities being excluded from the intervention (five non-START-Play behaviors). These 11 behaviors formed the basis of the fidelity measure to measure accuracy of delivery of the START-Play intervention and differentiate START-Play from uEI.

Step 2: establishing a measurement system

The START-Play fidelity measure consisted of two types of measurement: quantifiable therapeutic behavior (Table S1) and qualitative scoring via a Likert scale (Table S1). Individual behaviors were coded for each minute (i.e. if a behavior was observed any time during the interval, it was coded as '1'; if not observed, it was coded as '0'). Codes were not mutually exclusive and multiple codes could be documented in each 1-minute interval. The numbers of START-Play behaviors and non-START-Play behaviors occurring at least once in each minute were summed. Quality of the intervention was scored at the end of each 10-minute segment of the intervention session. Four rating scales with scores ranging from 1 (low) to 4 (high) were used to assess the overall quality of intervention. The first two scales were gestalt ratings, indicating the therapist's overall quality in providing START-Play intervention. The third and fourth scales were also gestalt ratings, indicating overall parental level of interest and engagement with the therapist and their child.

For each session, fidelity variables were calculated as follows. (1) Adherence to the START-Play intervention: the rate at which each START-Play behavior was used during a session (e.g. the number of minutes in which the behavior was used divided by total minutes of the session). (2) Dosage: total minutes of session, the number of minutes in which any START-Play behavior was used (START-Play dosage), and number of minutes in which any non-START-Play behavior was used (non-START-Play dosage). (3) Quality of intervention delivery: for both scales of therapists' overall level of effectiveness, Likert scale ratings for each 10-minute interval were averaged. (4) Participants' responsiveness: for parental level of interest and engagement scales, ratings for each 10-minute interval were averaged. (5) Program differentiation: the average of adherence, dosage, quality of intervention, and responsiveness were compared between groups.

Criteria for adequate adherence of the START-Play therapists was determined by scoring preliminary study intervention sessions that START-Play researchers considered to be good examples of the START-Play intervention. The criterion for each behavior and each summary variable was set a priori. These definitions and criteria were shared with the therapists during training.

Step 3: evaluating reliability evidence

Two coders (MA and AK) and two site principal investigators (RH and SD) of the START-Play research team worked together to clarify coding definitions and establish interrater reliability for all fidelity measure variables. Both coders had clinical and research experience in pediatric rehabilitation as a physical or occupational therapist and were not involved in steps 1 and 2. Reliability training involved didactic instruction by the principal investigators', coding of the same intervention videos, and comparison of coding results. The team met in person or via videoconference to discuss START-Play/non-START-Play behaviors on which the coders disagreed; they also refined the coding protocol and added examples of behaviors to clarify definitions in the protocol. While establishing reliability, some behaviors were coded in pairs to improve interrater reliability. For example, 'provides information on cognitive and motor interaction' and 'START-Play brainstorming' were paired for coding. If either of the two behaviors was observed during an interval, it was coded as '1', indicating presence. This iterative process of training and clarification continued until the two coders and one principal investigator reached acceptable agreement (≥80%) on four videos when comparing individual coding of all behavioral categories to be quantified. After reaching high agreement on

four consecutive videos, 20 videos (20% of the overall START-Play fidelity data; 10 START-Play and 10 uEI) were coded by both coders.[15] Intraclass correlation coefficients (ICCs) were calculated for each fidelity variable to examine interrater reliability evidence of START-Play fidelity scores using R-4.0.2 and the irr 0.84.1 package.[16] ICCs were calculated to quantify absolute agreement for individual ratings while treating participants as random effects and raters as fixed effects.[17]

Phase II: adherence and program differentiation using the fidelity measure

The second phase aimed to apply the START-Play fidelity measure to describe adherence to START-Play and quantify differences between START-Play and uEI services (program differentiation). The relationship between therapists' years of clinical experience and therapists' adherence to START-Play, and the relationship between therapists' adherence to START-Play and participant responsiveness were examined. The design of this phase was embedded in the overall comparison trial, comparing early intervention therapy sessions and START-Play sessions on the fidelity variables, as well as correlations between specific components of therapist characteristics and fidelity variables.

A total of 103 videos (64 START-Play and 39 uEI), with one video per child whose parent and therapist both consented in writing, were analyzed. Two coders with established interrater reliability and blind to group assignment coded the videos of the intervention sessions. The intervention videos were randomly assigned to either coder for independent coding.

Statistical analysis

Descriptive statistics for fidelity variables were calculated for the START-Play and uEI groups. To determine if the START-Play intervention was implemented as designed, variables on adherence and quality of intervention delivery in the START-Play group were compared with criteria for good fidelity (see Table S2, online supporting information, for criterion levels). General linear mixed-effects modelling, accounting for variance between coders and between videos, was used to describe differences between START-Play and uEI. Finally, to evaluate the relationships between intervention fidelity and participant responsiveness, and between therapists' length of clinical experience and their fidelity to START-Play intervention, bivariate correlation analyses, which accounted for data nesting within coders and videos, were conducted (Appendix S3, online supporting information). Data analyses were conducted with an alpha of 0.05 using R-4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and the lmerTest package 3.1.[18]

RESULTS

Phase I

The ICC values for all fidelity scores are shown in Table S2. All fidelity scores except two had good to excellent interrater reliability evidence with ICCs ranging from 0.77 to 0.95.[15] 'Greater assistance than needed' showed an acceptable level of reliability (ICC 0.73), whereas 'flexible and not rigid' had poor reliability. We further investigated 'flexible and not rigid' and found that when the ICC was computed separately by group, reliability evidence was good in the START-Play group (ICC 0.76) but poor in the uEI group. The overall high reliability findings supported the use of the fidelity measure in the START-Play clinical trial. In Phase II, we used the 'flexible and not rigid' variable for 'adherence' in the START-Play group but did not use it for program differentiation.

Phase II

Adherence to START-Play intervention

The START-Play therapists met the criteria for acceptable fidelity of the planned intervention (Table S2). Regarding adherence to START-Play behaviors, three of four behaviors met the criteria. Throughout the sessions, the therapists provided the child with 'cognitive opportunities' (mean rate [SD] 0.80 [0.17]) and showed 'flexible and not rigid' behavior (e.g. allowed the child to self-initiate movement and problem-solve) (mean rate 0.93 [0.11]). At a mean rate of 0.27 (0.22), the therapists provided information on cognitive and motor interaction or worked with the parent to brainstorm strategies for practicing START-Play concepts within the daily routines of the family. This was a slightly lower rate than the criterion (≥0.3). At a mean rate of 0.47 (0.25), therapists either encouraged parent engagement (e.g. providing information, brainstorming, encouraging parent to lead activities) or parents actively engaged in intervention. In terms of dosage, START-Play therapy sessions lasted 49.6 minutes on average. At least one START-Play behavior was implemented during an average of 48.1 minutes (96.8% of session length). For quality of intervention delivery and participant responsiveness, all quality ratings met the criteria (≥3 on a scale of 1–4). The findings indicated that START-Play therapists implemented key components of the intervention as intended, and the parents in the START-Play group engaged in a bidirectional discussion with the therapist, and interacted and provided opportunities for their child throughout most of the session.

Differentiation between START-Play and uEI services

Comparison analyses revealed significant differences between START-Play and uEI on all fidelity variables (Table S2; see Appendix S3 for detailed output). Regarding individual behavior use, rates of START-Play behaviors were higher in the START-Play group ($p<0.001$ for all behaviors) and rates of non-START-Play behaviors were higher in the uEI group ($p<0.001$ for all behaviors). In the uEI group, 'intervention activities which are not START-Play related' showed the highest rate of use (0.85 [0.18]). Therapists in the uEI group provided the child with 'greater motor assistance than needed' (0.35 [0.20]) and showed 'rigid adherence to correct way of moving' (0.51

[0.23]) at a much higher rate than therapists in the START-Play group (0.11 [0.14] and 0.09 [0.12] respectively). At a rate of 0.48, therapists in the uEI group provided information about the child's development other than the relationship between motor and cognitive skills or did brainstorming strategies that were not START-Play related. Therapists in the START-Play group implemented an average of 1.5 START-Play behaviors and an average of 0.3 non-START-Play behaviors during each minute of intervention. In the uEI group, for each minute of intervention, an average of 0.6 START-Play behaviors and an average of 1.5 non-START-Play behaviors were used.

In terms of dosage, therapy sessions lasted about 9 minutes longer in the START-Play group compared to the uEI group. The sum of minutes in which any START-Play behaviors were used (START-Play dosage) was larger in the START-Play group ($p<0.001$); the sum of minutes in which any non-START-Play behaviors were used (non-START-Play dosage) was larger in the uEI group ($p<0.001$).

For quality of intervention delivery and participant responsiveness, therapists in the START-Play group had higher ratings of overall effectiveness in providing a 'just right' challenge in sitting and reaching for the advancement of child cognition ($p<0.001$), as well as in initiating and encouraging parental interest and engagement ($p<0.001$) than those in the uEI group. Parental levels of engagement with the therapist ($p<0.05$) and their child ($p<0.001$) were also rated higher in the START-Play group.

Table 1: Bivariate correlations of adherence, quality of intervention delivery and participant responsiveness in START-Play group

	Parent interest/ engagement with therapist	Parent interest/ engagement with child
START-Play behavior use (adherence)	Correlation estimate (p)	
Cognitive opportunity	0.046 (0.685)	0.271 (**0.024**)
Flexible and not rigid	0.054 (0.664)	0.255 (0.052)
Provides information on cognitive and motor interaction OR START-Play brainstorm	0.42 (<**0.001**)	0.255 (0.048)
Encourages parent lead activities OR parent provides intervention	0.419 (<**0.001**)	0.793 (<**0.001**)
Average number of START-Play behaviors used during each minute	0.367 (**0.002**)	0.423 (**0.001**)
Quality of intervention delivery		
Therapist's level of effectiveness in using just right challenge in sitting and reaching to advance cognition	0.113 (0.399)	0.229 (0.086)
Therapist's level of effectiveness in initiating parental interest/engagement	0.866 (<**0.001**)	0.709 (<**0.001**)

Bold type indicates statistical significance. START-Play, Sitting Together and Reaching to Play.

Relationship between therapists' adherence to START-Play intervention, participant responsiveness, and therapists' clinical experience

Overall, positive significant correlations were indicated between adherence to START-Play intervention and participant responsiveness (e.g. parent engagement with the therapist and child) (Table 1). There was a moderate positive relationship among 'more START-Play behaviors per minute' and 'more minutes of therapists encouraging parental lead activity', and a 'higher level of parental engagement with the therapist'. There was a strong positive correlation between 'therapist encourages parent lead activities' and 'parent engagement with the child' throughout the intervention session. There was also a statistically significant relationship and strong positive correlation between 'therapists' effectiveness in initiating parental engagement' and 'parental level of engagement with the therapist and child'. Relationships between therapists' length of experience in early intervention and START-Play intervention fidelity (adherence to START-Play behaviors and quality of intervention delivery) were not significant ($p>0.05$).

DISCUSSION

Recently, fidelity measures have been developed and used in rehabilitation research, yet attention to measuring fidelity across the five dimensions and focusing on program differentiation has been lacking.[19–21] Monitoring of both the experimental and comparison groups, and comparing for key differences, is vital to optimize rehabilitation research. We described the development of a multidimensional, intervention-specific fidelity measure (Phase I) and the utilization of the measure to determine adherence and program differentiation (Phase II). By using the fidelity measure for both groups in a randomized controlled trial, the unique components of the START-Play intervention were clearly defined and differentiated from the intervention implemented by therapists in the uEI group.

Phase I

We developed the multidimensional fidelity measure following a three-step process and found evidence of reliability between coders after multiple iterations of training, analysis, and discussion. While establishing interrater reliability, refinements to the definitions of key components and the measurement system occurred. Although the exact time was not computed, training and practice for coding using the START-Play fidelity measure took more time than other fidelity measures (16 hours), as a rough estimation.[20] This may be partly because we used a complex measurement system which involved the rate of occurrence of each component rather than using a single score or checklist for present or absent.

Phase II

The START-Play fidelity measure was found to reliably identify the unique features and verify implementation adherence of the START-Play intervention. The fidelity process also distinguished the START-Play intervention from uEI. Therapists in the START-Play group used START-Play behaviors (particularly 'cognitive opportunity') with much higher rates, compared with the uEI therapists. The START-Play therapists kept cognition at the center of the task and provided the child with opportunities for practicing cognitive skills. The START-Play therapists were more effective in using the 'just right challenge' in sitting and reaching to advance the child's skills. Therapists in the START-Play group demonstrated non-START-Play behaviors (particularly 'rigid adherence to correct way of moving') at a much lower rate, which means they allowed the child time to solve their own motor problems and did not prohibit the child's self-initiated movements, consistent with the training and theoretical model.

Behaviors on certain aspects of family-centered services (e.g. sharing information and collaboration between therapists and parents) were observed in both groups, but the content was slightly different. Although information sharing or brainstorming was observed with similar rates of use (approximately 40% of the session) in both groups, the topics discussed in the START-Play group were primarily cognitive and motor interactions, but in the uEI group, motor-cognitive linkage was barely discussed. Instead, topics that related to children's development (such as motor and sensory functions, feeding, activities of daily living, adaptive equipment) were discussed. Although less frequent than in the START-Play group, the uEI group performed some START-Play behaviors such as 'encourages parent lead activities or parent provides intervention' at a rate that exceeded the criterion for adequate adherence of the START-Play therapist ($\geq$0.3). Behaviors that encourage parent engagement in intervention activities with their children may be a common feature of early intervention services. START-Play therapists also spent more time in initiating parental interest and engagement in the intervention than therapists in the uEI group. Finally, parents in the START-Play group were rated as demonstrating a greater level of interest and engagement with the therapist and with their child during the intervention, compared to parents in the uEI group.

Bivariate correlation analyses showed interesting findings. The therapists' length of experience in early intervention services was not significantly related to their adherence to the START-Play intervention or with the quality of intervention. This finding differs from other research showing positive significant correlations between quality of intervention and professionals' length of experience in practice.[22] One possible explanation is that START-Play therapists had a higher level of education (ranging from clinical masters to doctorate, median: clinical doctorate) compared to professionals who participated in the Knoche study (ranging from high school diploma to graduate degree, median: 4-year college degree).[22] This has clinical implications in that therapists who are educated to be an early intervention provider and properly trained in the START-Play intervention can implement this new intervention. In general, therapists' adherence to START-Play behaviors and quality of intervention delivery were significantly related to participant responsiveness. Interestingly, but understandably, START-Play behaviors that were correlated with participant responsiveness (i.e. parent interest/engagement) all involved interactions or communication between therapists and parents (e.g. providing information, brainstorming, encouraging parents to lead activities), rather than a focus on the child (e.g. cognitive opportunity). We believe that parental levels of interest/engagement influence intervention outcomes, however, the current study was not designed to study the effects of fidelity on intervention outcomes.

This research is novel in that it involves measurement of all five dimensions of fidelity and differentiates START-Play from other early intervention services. The process described can serve as a model for fidelity measurement development and implementation in intervention research. As opposed to generic fidelity measures that evaluate general attributes and strategies that are commonly used across similar interventions, the START-Play fidelity measure is an intervention-specific measure. It can be used for replicating efficacy or effectiveness studies on the START-Play intervention. The fidelity measure can also be used by therapists to examine their delivery of the key components of the START-Play intervention. Future work will include analysis of the relationship between fidelity outcomes and intervention outcomes, to identify the active elements of the START-Play intervention. This information can help to determine components that are essential for successful intervention and that make START-Play distinct from other early intervention services.

Limitations

One item on the fidelity measure ('flexible and not rigid') did not reach acceptable reliability evidence when used within the early intervention services, therefore, further study is needed to refine this item. Although the coders were blinded to group assignment, it is possible that they discovered unique features of START-Play intervention from watching videos. Consequently, they may have guessed which group a child belongs to, and this could have led to bias.

Thirty-five out of 65 therapists in the uEI group (53.8%) consented to be videotaped. There may be differences in demographic characteristics or in early intervention approaches utilized between the therapists who consented to be videotaped and those who did not. However, this is unknown because the therapists who were not videotaped also did not consent to filling out the survey that provided the relevant information. In addition, we were unable to specifically define 'usual care' early intervention with our methodology; our methods did not

include gathering data to completely describe early intervention across the different regions of the country. We could only describe the differences displayed from the START-Play intervention. Thus, the process we used (not the developed instrument) is more likely to be useful in future studies, which might aim to compare specific approaches (e.g. Goals Activity Motor Enrichment, Baby Constraint-induced Movement Therapy).[23,24] This study involved cross-sectional data. Examining intervention fidelity at multiple time points (e.g. multiple sessions of each family, multiple performances from each therapist across several families) would increase the generalizability of the findings across settings and groups. Future research will include longitudinal data from a larger sample of families and therapists, to examine differences in fidelity across time as well as the relationship between fidelity and intervention outcomes.

CONCLUSION

Fidelity measurement is important for randomized controlled trials in order to document adherence to intervention protocols and differentiation from the control condition. Development of an appropriate fidelity measure and determination of reliability evidence takes time and effort. After development and evaluation of the START-Play fidelity measure, we were able to document adequate adherence to the protocol and differentiation of START-Play from the uEI. Our detailed fidelity measurement process provides evidence of the reliability of the START-Play intervention comparison and can serve as a model for other intervention studies.

ACKNOWLEDGEMENTS

This study was funded by the US Department of Education, Institute of Education Sciences, award # R324A150103, clinical trial # NCT02593825. The authors have stated that they had no interests that might be perceived as posing a conflict or bias.

SUPPORTING INFORMATION

The following additional material may be found online:

Appendix S1: Examples of START-Play and uEI.

Appendix S2: Demographic information of therapists in START-Play and uEI group.

Appendix S3: Detailed output of comparisons between START-Play and uEI on all fidelity variables.

Table S1: Behaviors coded/scored and operational definitions

Table S2: Interrater reliability and descriptive statistics for fidelity variables

REFERENCES

1. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev* 1998; **18**: 23–45.
2. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Educ Res* 2003; **18**: 237–56.
3. O'Donnell CL. Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Rev Educ Res* 2008; **78**: 33–84.
4. Gresham FM, Gansle KA, Noell GH. Treatment integrity in applied behavior analysis with children. *J Appl Behav Anal* 1993; **26**: 257–63.
5. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. *Clin Psychol Rev* 1991; **11**: 247–66.
6. An M, Dusing SC, Harbourne RT, Sheridan SM. What really works in intervention? Using fidelity measures to support optimal outcomes. *Phys Ther* 2020; **100**: 1–9.
7. Salamh P, Cook C, Reiman MP, Sheets C. Treatment effectiveness and fidelity of manual therapy to the knee: A systematic review and meta-analysis. *Musculoskeletal Care* 2017; **15**: 238–48.
8. Walker MF, Hoffmann TC, Brady MC, et al. Improving the development, monitoring and reporting of stroke rehabilitation research: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Neurorehabil Neural Repair* 2017; **31**: 877–84.
9. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014; **348**: g1687.
10. Vaughan-Graham J, Cott C, Wright FV. The Bobath (NDT) concept in adult neurological rehabilitation: what is the state of the knowledge? A scoping review. Part I: conceptual perspectives. *Disabil Rehabil* 2015; **37**: 1793–807.
11. Harbourne RT, Dusing SC, Lobo MA, et al. Sitting Together And Reaching To Play (START-Play): protocol for a multisite randomized controlled efficacy trial on intervention for infants with neuromotor disorders. *Phys Ther* 2018; **98**: 494–502.
12. Lobo MA, Harbourne RT, Dusing SC, McCoy SW. Grounding early intervention: physical therapy cannot just be about motor skills anymore. *Phys Ther* 2013; **93**: 94–103.
13. Mowbray CT, Holter MC, Teague GB, Bybee D. Fidelity criteria: Development, measurement, and validation. *Am J Eval* 2003; **24**: 315–40.
14. Sheridan SM, Marvin C, Knoche L, Edwards CP. Getting ready: Promoting school readiness through a relationship-based partnership model. *Early Childhood Services* 2008; **2**: 21–45.
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; **15**: 155–63.
16. Gamer M, Lemon J, Singh P. irr: Various coefficients of interrater reliability and agreement (0.84.1), 2019.
17. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; **1**: 30–46.
18. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw* 2017; **82**: 1–26.
19. Quinn L, Trubey R, Gobat N, et al. Development and delivery of a physical activity intervention for people with Huntington disease: facilitating translation to clinical practice. *J Neurol Phys Ther* 2016; **40**: 71–80.
20. Rezze BD, Law M, Eva K, Pollock N, Gorter JW. Development of a generic fidelity measure for rehabilitation intervention research for children with physical disabilities. *Dev Med Child Neurol* 2013; **55**: 737–44.
21. Toomey E, Matthews J, Guerin S, Hurley DA. Development of a feasible implementation fidelity protocol within a complex physical therapy-led self-management intervention. *Phys Ther* 2016; **96**: 1287–98.
22. Knoche LL, Sheridan SM, Edwards CP, Osborn AQ. Implementation of a relationship-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. *Early Childhood Research Quarterly* 2010; **25**: 299–313.
23. Morgan C, Novak I, Dale RC, Guzzetta A, Badawi N. Single blind randomised controlled trial of GAME (Goals Activity Motor Enrichment) in infants at high risk of cerebral palsy. *Res Dev Dis* 2016; **55**: 256–67.
24. Eliasson AC, Nordstrand L, Ek L, Lennartsson F, Sjöstrand L, Tedroff K, Krumlinde-Sundholm L. The effectiveness of baby-CIMT in infants younger than 12 months with clinical signs of unilateral-cerebral palsy; an explorative study with randomized design. *Res Dev Dis* 2018; **72**: 191–201.