The Pennsylvania State University

The Graduate School

ADVANCING MEASUREMENT TECHNOLOGY IN EDUCATIONAL INTERVENTION RESEARCH TO STUDY INDIVIDUAL, CONTEXTUAL, AND IMPLEMENTATION HETEROGENEITY

A Dissertation in

Human Development and Family Studies

by

Kyle D. Husmann

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2023

The dissertation of Kyle D. Husmann was reviewed and approved by the following:

Timothy Brick
Associate Professor of Human Development and Family Studies
Dissertation Advisor
Co-Chair of Committee

Zita Oravecz
Associate Professor of Human Development and Family Studies
Co-Chair of Committee

James DiPerna
Professor of Education

Karen Bierman
Professor of Psychology and Human Development and Family Studies

Charles Geier
Professor of Human Development and Family Studies
Director of Graduate Studies

ABSTRACT

For the past two decades, educational intervention research has rallied around the common goal of determining "what works in education" to inform education practice and policy (U.S. Department of Education, 2017). To study individual, contextual, and implementation processes and their interaction with educational interventions requires dramatically scaling up the amount of data we collect in educational research (Bryan et al., 2021). Existing technology has the potential to dramatically increase our data collection capacity without the cost and burden associated with traditional measurement approaches. In this dissertation I present two pilot studies that demonstrate tractable and immediately accessible ways that mobile technology can expand our ability to test theory about educational processes. In the first study, I present results from an Ecological Momentary Assessment (EMA) study that used a mobile app to collect parents' daily levels of parental self-efficacy for their child's home literacy activities. Although cross-sectional studies have shown relationships between parental self-efficacy and their involvement with their child's literacy activities in the home environment (e.g. Dulay et al., 2018; Tazouti & Jarlégan 2016; Giallo, 2013), the extent to which these relationships represent day-to-day relationships existing within parents is not known. I test competing predictions from social cognitive theory (Bandura, 2000) and resource allocation theory (Yeo & Neal, 2006) about the within-person dynamics of parental self-efficacy and their child's home literacy activities. In the second study, I demonstrate how mobile technology can augment materials in an existing intervention in order to capture low-burden, intensive longitudinal measures of implementation and implementation-related processes. I present a proof-of-concept mobile app that I developed to replace the paper forms of an existing parent-implemented literacy intervention, and I share results from a pilot test of its feasibility and acceptability. I demonstrate the app's ability to automatically collect dense longitudinal measures during the intervention with negligible burden to parents, children, and researchers. In analyses using the novel data I collected in the pilot, I test hypotheses about within-person relationships existing in parents' and children's emotional experiences during literacy activity sessions and their future engagement with the intervention. I conclude with a vision for the future of measurement and data collection in educational intervention research. I introduce Cattell's (1952) developmental "Data Box" framework and show how it can organize the types of meaningful differences existing in evaluations of educational interventions. Finally, I argue that expanding our measurement capability to study heterogeneity in all its forms is not only necessary for advancing our scientific knowledge about education, but also a moral imperative to ensure that applications of our scientific knowledge provide equitable benefits to the world.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply grateful to my dissertation committee for all their support and guidance throughout my research journey. Thank you to Dr. Timothy Brick and Dr. Zita Oravecz, my mentors in HDFS, for cultivating such a healthy, open, and collaborative research environment. Thank you to Dr. James DiPerna and Dr. Karen Bierman, for your roles in the TIES fellowship program and for being my bridge into the world of education research.

Thank you to Sharon Juenemann and Kathryn Stevenson from the Shadow Project for their invaluable collaboration in the design and implementation of the study I present in Chapter 2. It would not have been possible without their insight and experience. Thank you to Douglas Research Endowment & Hintz Fellowship for the funding to make this work possible.

I would like to express my deep gratitude to Kathryn Stevenson, the founder of Syllabee, for her invaluable collaboration in the study presented in Chapter 3. Her generous contribution of time and insightful guidance during the app development and implementation process have been instrumental. Additionally, I would like to extend my thanks to the Ford Foundation Scholarship for their funding, which made this work possible.

And of course, thank you to my friends and family, for all your support and encouragement during this chapter of my life.

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grant R305B090007 to The Pennsylvania State University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

"If that is what science at its best can be like, what kind of world does that leave us with? It looks like we are left with a world just like the one we actually live in: a world that contains within its boundaries countless small worlds where precise order and stability obtain, but a world that is itself big, wide and gloriously differentiated, where much that happens beyond the confines of the small worlds within it, happens haphazardly -- a result of the interaction of causes studied across the scientific disciplines, from fundamental particle physics to neuroscience to social psychology. This is a world in which irritability, generosity and social exclusion can affect what happens just as gravity and electromagnetic repulsion can."

- *Nancy Cartwright*

"All models are wrong, some are useful."

- *George Box*

Chapter 1

Introduction

For the past two decades, educational intervention research has rallied around the common goal of determining "what works in education" to inform education practice and policy (U.S. Department of Education, 2017). Despite the influx of high-quality causal evaluations of educational interventions, however, it has become clear that "what works in education" is not universally generalizable. Because intervention effects can vary widely due to heterogeneity in individuals, contexts, and implementations, questions about "what works" in general have given way to "what works, for whom, under what circumstances" (Gutiérrez & Penuel, 2014; Weiss et al., 2014). Recent reflections by major stakeholders on the future of educational intervention research have named "issues of heterogeneity" a "primary concern" for education intervention research moving forward into this next decade (Committee on the Future of Education Research, 2022).

To study individual, contextual, and implementation differences and their interaction with educational interventions, we must have the capacity to measure these differences (Bryan et al., 2021). This means dramatically scaling up the amount of data we collect in educational research, by increasing the number and type of measures we collect, the number of measurement occasions, and the number of individuals we sample. Traditionally, increasing the amount of data has come at the cost of additional participant burden and research expense. Unfortunately, when data collection is burdensome, invasive, and costly, we collect less data, thereby stunting our ability to study individual, contextual, and implementation differences. If we wish to study "issues of heterogeneity" in this next decade, it is imperative we also directly address these issues of data collection.

Existing technology has the potential to dramatically increase our data collection capacity without the cost and burden associated with traditional measurement approaches. In the following chapters I present two pilot studies that demonstrate tractable and immediately accessible ways that technology can expand our measurement capacity in educational intervention research. Although the technological approaches I present in these studies can be applied to any educational context (e.g., a classroom, a clinic, etc.), I focus on the measurement of home literacy practices and intervention, which is especially affected by the challenges and limitations of traditional measurement approaches.

In Chapter 2, I present results from an Ecological Momentary Assessment (EMA) study that used a mobile app to collect parents' daily levels of parental self-efficacy and their child's home literacy activities. Theoretical models of parent involvement suggest a process-level relationship between parental self-efficacy and their involvement with their child's literacy development (e.g., Hoover-Dempsey & Sandler, 1995). Cross-sectional studies have shown between-person relationships in parental self-efficacy and their children's home literacy activities (e.g., Dulay et al., 2018; Tazouti & Jarlégan 2016; Giallo, 2013), but their within-person, day-to-day dynamics have not been empirically investigated. Social cognitive theory (Bandura, 2000) and resource allocation theory (Yeo & Neal, 2006) provide competing predictions for the relationships in the day-to-day dynamics of parental self-efficacy and their children's home literacy activities. Using the daily measures of parental self-efficacy and children's home literacy activities collected by way of the mobile app, I tested these predictions.

In Chapter 3, I demonstrate how technology can be used to obtain measures of parent enactment and enactment-related processes in family literacy programs (FLPs) without the cost and burden associated with traditional measurement approaches. For FLPs, parents' regular enactment of literacy activities with their child is the "core engine" for the intervention's theory of change. Understanding the individual processes that support or derail parent enactment can help us design more contextually sensitive FLPs. Process models of parent involvement suggest parents' and children's affective experiences during literacy activities may impact their motivation to engage in future enactment sessions. I piloted a novel FLP implementation measurement approach to test these hypotheses. I developed a mobile app to replace the paper materials of an existing FLP and automatically collect dense longitudinal measures during the intervention with negligible burden to parents, children, and researchers. I tested the app's acceptability and usability by parents. With the pilot data I collected, I tested for evidence of the predicted within-person relationships existing in parents' and children's emotional experiences during literacy activity sessions and their future engagement with the intervention. Finally, I discuss possible extensions of this measurement approach.

I conclude with a vision for the future of measurement and data collection in educational intervention research. I discuss how empirically establishing the limits of generalizability of our theories and models will become even more necessary as technology enables more detailed and intensive measurement. I echo recent calls for the investment in shared research data infrastructure so that we can obtain the large datasets necessary for answering these questions. I introduce Cattell's (1952) developmental "Data Box" framework and show how it can organize the types of meaningful differences existing in evaluations of educational interventions. Finally, I argue that expanding our measurement capability to study heterogeneity in all its forms is not only necessary for advancing our scientific knowledge about education but is a moral imperative to ensure that applications of our scientific knowledge provide equitable benefits to the world.

References

Bandura, A. (2000). Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Bryan, C., Tipton, E., & Yeager, D. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01143-3

Cattell, R. B. (1952). The three basic factor-analytic research designs—Their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499–520. https://doi.org/10.1037/h0054245

Committee on the Future of Education Research at the Institute of Education Sciences in the U.S. Department of Education, Board on Science Education, Division of Behavioral and Social Sciences and Education, & National Academies of Sciences, Engineering, and Medicine. (2022). *The Future of Education Research at IES: Advancing an Equity-Oriented Science* (A. Gamoran & K. Dibner, Eds.; p. 26428). National Academies Press. https://doi.org/10.17226/26428

Dulay, K. M., Cheung, S. K., & McBride, C. (2018). Environmental correlates of early language and literacy in low- to middle-income Filipino families. *Contemporary Educational Psychology*, *53*, 45–56. https://doi.org/10.1016/j.cedpsych.2018.02.002

Giallo, R., Treyvaud, K., Cooklin, A., & Wade, C. (2013). Mothers' and fathers' involvement in home activities with their children: Psychosocial factors and the role of parental self-efficacy. *Early Child Development and Care*, *183*(3–4), 343–359. https://doi.org/10.1080/03004430.2012.711587

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to Practice as a Criterion for Rigor. *Educational Researcher*, *43*(1), 19–23. https://doi.org/10.3102/0013189X13520289

Tazouti, Y., & Jarlégan, A. (2019). The mediating effects of parental self-efficacy and parental involvement on the link between family socioeconomic status and children's academic achievement. *Journal of Family Studies*, *25*(3), 250–266. https://doi.org/10.1080/13229400.2016.1241185

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2017). What works clearinghouse standards handbook (Version 4.0). *Washington, DC: Institute of Educational Sciences Retrieved from Https://Ies. Ed. Gov/Ncee/Wwc/Handbooks*.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A Conceptual Framework for Studying the Sources of Variation in Program Effects. *Journal of Policy Analysis and Management*, *33*(3), 778–808. https://doi.org/10.1002/pam.21760

Yeo, G. B., & Neal, A. (2006). An examination of the dynamic relationship between self-efficacy and performance across levels of analysis and levels of specificity. *Journal of Applied Psychology*, *91*(5), 1088–1101. https://doi.org/10.1037/0021-9010.91.5.1088

Chapter 2

Day-to-Day Relationships in Parental Self-Efficacy and Home Literacy Practices

A significant proportion of a child's literacy development occurs outside of the classroom context, by way of parent-child interactions in the home environment (Weinberger 1996; Burgess et al., 2002). Parents affect their child's literacy and language trajectories starting from a very young age by way of a variety of mechanisms, ranging from daily verbal interactions with their child (Hoff, 2003; Rowe et al., 2005), to shared reading activities with their child that stimulate their early literacy development (Singh et al., 2022; Sénéchal et al., 1996). Even as children get older and gain more independence in their reading ability, parents can continue to play an active role in their child's literacy development by cultivating their child's interest and motivation around independent book reading (Silinskas et al., 2020; Boerma, 2018).

Observational studies suggest parents' involvement with their children in literacy-related activities mediates relationships between family characteristics (e.g., SES) and children's academic outcomes. Findings have generally indicated that parents with less education and available resources typically engage their children in fewer literacy-related activities, which in turn is associated with lower levels of childrens' literacy ability (Luo et al., 2021; Burris et al., 2019; Raikes et al., 2006; Rodriguez et al., 2009; van Steensel 2006). Although these results are useful for identifying which families might benefit from extra support, more work is needed to identify and understand the underlying processes that may be driving these observed differences between families. Process models of human behavior seek to explain individual differences as arising from complex interactions between individuals and their environment across time, rather than as simple consequences of "intrinsic" characteristics (Bronfenbrenner & Morris 2006). A process-level perspective is particularly useful for studying differences in parental involvement because it explicitly acknowledges the role of parents' changing experiences across time in shaping their decisions. Understanding how individual experiences relate to differences in parents' level of involvement with their child's literacy development can help us develop more contextually sensitive family literacy interventions that rely on forms of parent engagement with their children as their primary mechanism of action (Baker, 2013; Anderson et al., 2010; van Steensel et al., 2014; Mol et al., 2008).

Process Models of Parental Involvement

The Hoover-Dempsey and Sandler's (1995, 1997) model of parental involvement in education and its later iterations (Green et al., 2007; Walker et al., 2005; Anderson & Minke et al., 2007) give a process-level account of how parental experiences might impact their levels of involvement with their child's literacy development. Although Hoover-Dempsey and Sandler's model began as a five-layered chain of processes connecting parental characteristics to children's educational outcomes through parental involvement, later versions reorganized these layers into three key determinants of parental involvement, understood as consequences of parents' individual, relational, and contextual experiences: parents' motivational beliefs, parents' perceptions of invitations for involvement, and parents perceptions of their life context (Green et al., 2007; Walker et al., 2005).

Although these models were originally designed to explain parents' general levels of involvement with their child's education both at home and at school, they are still applicable to the more specific case of parental involvement with their child's home literacy activities. For example: on a day in which the parent experienced stress at work, had a conflict with their child earlier that day, and did not sleep well the night before, the Hoover-Dempsey and Sandler model predicts lower levels of parental involvement in their child's literacy activities by way of diminished levels of parent motivation, fewer invitations from their child to engage in literacy activities, and a decrease in the parent's perception of the amount of time they have available to spend with their child.

The Hoover-Dempsey & Sandler model of parental involvement and its variations draw heavily on social-cognitive theories of motivation and behavior change to explain how parent experiences will influence their parenting decisions and involvement behaviors (Schunk & DiBenedetto, 2020; Bandura, 1977; Bandura, 2000). Social-cognitive theory hypothesizes that experiences change behaviors by way of shaping beliefs and perceptions as well as affecting levels of motivation. Central to social cognitive theory is the construct of self-efficacy, defined as an individual's beliefs about their ability to exercise self-agency: that their actions can produce desired effects (Bandura, 2000). In the Hoover-Dempsey & Sandler model and its related variations, parents' levels of motivation to be involved with their child's education are determined by their beliefs about their role in their child's education (Lynch et al., 2007; Bingham, 2007; Weigel et al., 2006), as well as their parental self-efficacy in engaging in the parenting and educational activities that characterize parental involvement, such as reading with their child (Yeo et al., 2014), helping their child with their math homework (Wu et al., 2022), or volunteering at their child's school (Castro et al., 2004).

Parental Self-Efficacy

Parental self-efficacy, more broadly defined as parent's perceptions of their parenting skill and ability to be a good parent, has been shown to simultaneously shape parent behavior and act as an indicator for contextual stressors that may be impacting parents' perceived abilities (Jones & Prinz, 2005). Social cognitive theory predicts that parents' feelings of parental self-efficacy will be increased when they have "mastery experiences" of their parenting abilities: occasions characterized by successful and enjoyable interactions with their child (Schwarzer & Luszczynska 2008). These feelings of parental self-efficacy are thought to then influence their parenting behaviors by making them more likely to initiate other positive parenting interactions with their child, which could extend to their involvement with their child's home literacy activities (Coleman & Karraker, 1998).

Empirical evidence has demonstrated a parent-level relationship between parents' parental self-efficacy and their levels of involvement with their child's home literacy activities. Most notably, in a study on parent involvement in home activities with their children, Giallo (2013) found that parental self-efficacy mediated relationships between parent-wellbeing, child temperament, and parents' involvement in home activities with their child (which included shared literacy activities). Specifically, parents who reported being more stressed, depressed, or had a child with more behavioral challenges reported less parental self-efficacy, which predicted lower frequencies of activity engagement with their child. Giallo hypothesized that these results could be an indication that parents who were not feeling well, or had children that exhibited more behavioral difficulties, might also be having a harder time enjoying their interactions with their child, making them feel like less successful parents, and become less likely to engage in these parent-child activities. In the same vein, Dulay et al. (2018) and Tazouti and Jarlégan (2016) found positive correlations between parental self-efficacy and levels of parents' involvement with their child's (including shared literacy) home literacy activities. Additionally, though their samples represented very different populations and cultures, they both found that parental self-efficacy and parent involvement mediated the relationship between SES and children's early literacy and language skills, highlighting both constructs as potential targets for intervention.

Although these findings support the predicted relationship between levels of parental self-efficacy and "typical" levels of parental involvement with their child's literacy activities, more work is necessary in order to determine the extent to which these findings represent relationships in the underlying processes these constructs represent. As processes, parental self-efficacy and parent interaction with their children exhibit meaningful fluctuations across time arising from reciprocal interactions between parents, children, and their environments (Bronfenbrenner & Morris 2006). Social cognitive theory, as a process-level theory, makes predictions about the relationships we should expect to observe in parents' fluctuations of self-efficacy and parent interaction with their children across time as they are shaped by their individual, relational, and contextual experiences (Yeo & Neal, 2013). Extant work relating parental self-efficacy and involvement tests hypotheses about the differences we might expect to find between individuals due to the distal effects of these processes aggregating over time, but cannot untangle the extent to which the observed individual differences are related to an interaction of the hypothesized processes across time, or by some other unmeasured relationship (Chen et al., 2005).

Directly testing theory about the processes driving between-person relationships in parental self-efficacy and involvement requires frequent observations of these constructs as they change across time within parents (Nesselroade & Ram, 2004). By relating the fluctuations in parental self-efficacy and involvement within parents across time, we can directly test social cognitive theory's more proximal predictions about relationships in these constructs, in addition to their hypothesized aggregate, between-person distal effects (Chen et al., 2000). Where previous work has found that parents who have higher levels of parental self-efficacy also have higher levels of parental involvement in their child's literacy activities, obtaining frequent measures of self-efficacy and parental involvement makes it possible to test if this same relationship exists within an individual parent's experience across different days. Theoretical models of parental involvement rooted in social cognitive theory would predict that on days that a given parent experiences more success in engaging their child in literacy activities, their feelings of success that day will contribute to higher levels of parental self-efficacy that day. And then, social cognitive theory predicts that these higher levels of parental self-efficacy will translate into increased levels of involvement with their child's literacy activities in the proximal future (e.g. the next day). With daily measures of parents' parental self-efficacy and their child's literacy activities, along with statistical analyses that untangle day-to-day relationships from their average trends across parents (e.g. Wang & Maxwell, 2015), we can directly test these hypotheses.

The Dynamics of Self-Efficacy

Although process models based on social-cognitive theories of parental involvement would predict a positive day-to-day association between parental self-efficacy and their levels of involvement with their child's literacy development, other theories of motivation and decision-making make the opposite within-person prediction: a negative day-to-day relationship between parental self-efficacy and levels of involvement. Previous work studying the dynamics of self-efficacy across time has suggested resource allocation theory may play a role in the between-person relationships of self-efficacy and individuals' decision-making (Yeo & Neal, 2006; Yeo & Neal 2013; Vancouver et al., 2001; Beck & Schmidt 2012). Like social cognitive theory, resource allocation theory explains human decision-making by way of an individual's motivations, beliefs, and perceptions, but unlike social cognitive theory, which focuses primarily on the positive role of self-efficacy in motivation, resource allocation theory also considers how finite levels of cognitive and attentional resources can influence behavior (Kahneman, 1973; Kanfer & Ackerman, 1989).

From a resource allocation theoretical perspective, an individual's self-efficacy on a task is an indicator of how difficult they think it will be, where the less self-efficacy someone has for a task, the more challenging they perceive it to be. When the individual is juggling competing goals with finite resources, increases in an individual's self-efficacy on a task are expected to relate to decreases in the amount of attention they expect it to require in the future, and therefore be associated with decreases in the future levels of attention they allocate to that task (Ballard et al., 2016). Like social cognitive theory, resource allocation theory predicts a positive between-person relationship between self-efficacy and task performance, because self-efficacy is expected to correlate with ability. But unlike social cognitive theory, resource allocation theory predicts a negative within-person relationship between self-efficacy and task performance. This hypothesis is supported by empirical findings by Yeo and Neal (2006) in a study on dynamic relationships between self-efficacy and performance on an air traffic control task. Although Yeo and Neal found a positive between-person relationship in self-efficacy and performance as both social cognitive theory and resource allocation theory would predict, they also measured a negative within-person relationship in these constructs, in support of resource allocation theory. Similarly, Vancouver et al. (2001) found negative within-person relationships between undergraduates' self-efficacy and their within-person changes in academic performance.

Applied to parent involvement in home literacy activities, resource allocation theory suggests that a parent's increases in self-efficacy will be associated with a decreased estimate of the amount of energy or effort they expect involvement to require in the future. A parent's involvement with their child's literacy activities is not their only parenting goal; so when they perceive that their involvement will require less effort, they may choose to spend their attentional "savings" on other tasks or priorities. When a parent's experiences with their child makes their parenting activities feel easier, or like they are doing a good job, their increased self-efficacy may actually cause them to deallocate future attentional and cognitive resources to engagement with their child.

That parents may choose to increase their involvement in the face of increased levels of difficulty or challenge is not without some empirical support in parental involvement research. In longitudinal studies of changing parental involvement at the year-to-year level, both Sénéchal and LeFevre (2014) and Silinskas et al. (2020) found that a child's reading difficulties and challenges at the end of preschool predicted increases in their parent's involvement in their literacy activities at the end of grade 1. These results can be interpreted in support of a resource allocation model of parental decision-making in which within-person increases in a parent's perception of challenge is positively associated with increases of their involvement in their child's literacy activities. The key difference, of course, between the findings of these studies and potential relationships between parental self-efficacy and involvement, is that the hypothesized mechanism of assessed "challenge" was based parents' perceptions of their child's abilities, rather than parents' perceptions of their own abilities, which may very well exhibit different dynamics.

Present Study

In the present study, we tested the hypothesis that a parent's parental self-efficacy and their involvement with their child's literacy activities are related by way of their underlying processes, as predicted by social cognitive theory and resource allocation theory. To do this, we adopted an Ecological Momentary Assessment (EMA; Shiffman et al., 2008; Carson et al., 2010) study design consisting of daily measures of parental self-efficacy and involvement with their child's literacy activities. In a 14-day measurement burst, 16 parents used a mobile app to answer daily questions about their parental self-efficacy and home literacy activities along with other measures of their contextual state. Collecting these intensive longitudinal data enabled us to test for systematic relationships in the dynamics of these constructs across time using statistical analyses that untangled day-to-day relationships from their average trends across parents. We test the following hypotheses:

Hypothesis 1: Children of parents with higher levels of parental self-efficacy also engage in higher average levels of daily literacy activities. This hypothesis describes a between-person relationship between parental self-efficacy and their levels of involvement in their child's literacy activities. Based on previous studies, and the predictions of both social cognitive theory and resource allocation theory, we expect this relationship to exist and be positive.

Hypothesis 2: On days parents report that their children were engaged in higher levels of literacy activities, they will also report higher levels of parental self-efficacy. This hypothesis describes a retrospective, within-person relationship between parental self-efficacy and their child's literacy activities. Finding a significant, within-person relationship here provides evidence that parental self-efficacy and their child's literacy activities have a process-level relationship. Both social cognitive theory and resource allocation theory predict a positive relationship here as well: that a parent's success will proximally boost their feelings of parental self-efficacy.

Hypothesis 3: Parents' level of parental self-efficacy on one day will be systematically related to levels of their child's literacy activities the next day. This final hypothesis describes a prospective, within-person relationship between parental self-efficacy and their child's literacy activities. Here, social cognitive theory and resource allocation theory make opposite predictions. Social cognitive theory predicts this relationship will be positive, because a parent's feelings of success one day will increase their likelihood to engage with their child the next day. Resource allocation theory predicts a negative relationship, attributed to parents' allocation of less attention and cognitive resources to an activity when it is perceived as less challenging. After testing these hypotheses, we discuss the implications of our findings to our understanding of the processes underlying parents' parental self-efficacy and their involvement in their child's home literacy activities, and possible applications to home literacy interventions.

Method

Approach

This study was run in partnership with The Shadow Project, a 501c3 nonprofit organization that exists to support children with special needs by providing parents, special education teachers, and children with supplementary programs, services, and materials. Their virtual goal setting program, which began in the summer of 2021, served as the recruitment pool for parents in this study (see Participants, below). The study took place in the weeks following the last day of school and prior to the start of the Shadow Project's virtual goal setting program.

In order to answer our research questions about differences between parent days, we needed to obtain parent reports of day-to-day levels of home literacy activities, parenting self-efficacy, and time resources. To collect these measures, we adopted an Ecological Momentary Assessment approach (EMA; Carson et al., 2010) consisting of brief daily surveys delivered to parents through Wear-IT, mobile app designed for conducting EMA studies (Brick et al., 2020).

Participants

16 parents of children in grades K-5 were recruited from the Shadow Project's network of approximately 30 K-8 public schools in Multnomah, Marion, and Yamhill counties in northwest Oregon. Recruitment for this study piggybacked on the Shadow Project's outreach for their Summer Virtual Goal-Setting Program for children with special needs. Parents signing up to participate in the Goal Setting Program were asked if they were also interested in participating in this study, and then received follow-up communication to screen for eligibility in this study and guide them through the installation of the Wear-IT mobile app. Parents provided informed consent through the Wear-IT mobile app.

Parents were **included** in the study if the parent:

- Self-reported that they have a child in grade K-5
- Consented to participate in this study and receive daily surveys about their parenting and home literacy environment.
- Spoke English fluently.
- Owned a mobile device (IOS or Android) that supported the Wear-IT mobile app.

Parents were **excluded** from this study if the parent:

- Was unwilling to install the Wear-IT app on their mobile phone.
- Reported that they planned to travel or take a vacation anytime during the two weeks of the study duration.

Measures

We adapted existing measures of parental resources, parenting self-efficacy and home literacy environment and administered them in a daily diary format. We also administered the original versions of these measures in a baseline, retrospective survey to verify the internal validity of our adapted day-to-day versions of these measures. Demographic information, like parental education, and child grade were also collected in the baseline survey.

Home Literacy Activities

In line with previous work on conceptualizations of the home literacy environment (e.g. Burgess et al., 2002), we adapted measures of home literacy activities from the Stony Brook Family Reading Survey (Payne et al., 1994). In the baseline assessment, we asked parents "How often do you read with your child" and "How often does your child read to themselves," which parents answered on a three-point Likert scale (1 = weekly or less, 2 = several times a week, or 3 = daily).

On the daily survey, before asking parents about their child's literacy activities that day, we first first asked if they were with their child at all in the previous 24 hours. If they indicated they were not with their child, we skipped the questions about their child's literacy activities and recorded their responses as missing.

After we confirmed that parents were with their child that day, we then asked parents if they spent time reading with their child that day. Then if they answered affirmatively, parents had the opportunity to estimate the amount of time they spent reading with their child across types of literary media (e.g. digital books, printed books, picture books, etc.). Parents estimated the amount of time for each category in 10-minute increments up to 60 minutes (with an option to mark "more than 60 minutes", which we coded as 70 minutes). In our analyses, to determine the total time a parent spent reading with their child that day, we took the sum of their responses across all of the possible literary media categories (including "other").

Daily measures of child independent reading habits were obtained similarly. First, we asked parents if their child read independently at all that day. Then, if they indicated their child had read independently that day, we asked parents to indicate the amount of time their child spent reading different types of literary media, using a scale of 10-minute increments up to 60 minutes, as before (again with an option to mark "more than 60 minutes"). Finally, we calculated the total amount of independent reading for that day by taking the sum of their time estimates across the literary media categories.

Children's Reading Enjoyment and Struggle

In the baseline survey, we included two items asking parents about their child's enjoyment of reading and how much they struggled with reading. Parents rated their child's reading enjoyment on a visual analog scale (Chang & Little, 2018) in the Wear-IT app from "Not at all" to "A lot" and responses were coded from 1 to 5. Similarly, parents rated their child's reading struggle on a scale from "No struggle" to "Lots of struggle" using a visual analog scale with responses coded from 1 to 5. These variables were only obtained in the baseline survey at the person-level and were not included in the daily questioning. These responses were used as covariates in our analyses.

Parenting Self-Efficacy

Parenting self-efficacy was measured at baseline using Dumka et al.'s (1996) Parenting Self-Agency Measure (PSAM). This measure consists of 8 items asking parents to rate their feelings about their parenting confidence and ability (e.g., "I know I am doing a good job as a parent" and "I feel sure of myself as a parent"). Responses to the PSAM are made using a 7-point discrete Likert scale (1 = rarely to 7 = always). In the Wear-IT app we used a visual analog scale (Chang & Little, 2018) that coded responses to the same anchors. In previous work the PSAM has been shown to be a robust measure, exhibiting acceptable levels of reliability (Chronbach's alpha > 0.7) even in cross-cultural samples (Dumka et al., 1996). The baseline parenting self-efficacy score used in our analyses was obtained using the mean of the responses to the 8 items.

To minimize participant burden in the daily responses, we selected only two out of the eight items from the PSAM, and reframed the statements to specifically reference the previous 24 hours of the parent's experience: "I know I did a good job as a parent in the last 24 hours" and "I felt sure of myself in my parenting in the last 24 hours". Daily responses to these questions were also made using a visual analog scale that placed items on a continuum from 0 to 5 (0 = strongly disagree to 5 = strongly agree). Responses from these 2 items were averaged to obtain the parent's daily parenting self-efficacy score used in our analyses.

Parental Time Resources

To assess parental time resources at baseline, we used the Family Resource Scale-Revised (FRS-R; Van Horn et al., 2001). The FRS-R asks parents to rate the extent to which they have adequate resources to meet the needs of their family in four domains: basic needs, money, time for self, and time for family. In previous psychometric evaluations, the FRS-R has exhibited acceptable levels of reliability (Cronbach's alpha > 0.7) and demonstrated external validity in relationships to other common measures of family resources (Van Horn et al., 2001). Responses are rated on a 5-point discrete Likert scale (1 = not at all adequate and 5 = almost always adequate). In the Wear-IT app we used a visual analog scale (Chang & Little, 2018) that coded responses to the same anchors. We did not include an option to select "does not apply", as with the original FRS-R. For this study, because we were specifically interested in measuring parents' time resources, we formed our baseline parental time resources measure by taking the mean of the 8 items belonging to the "time for self" and "time for family" subscales, which included items ranging from "time to be with children" and "time to be by self".

As with the PSAM, we chose a subset of the items to minimize participant burden for the daily measurement. We chose only items 6 from the "time for self" and "time for family" subscales. We reframed the statements to specifically reference the previous 24 hours, and adjusted the wording on a few of the items to clarify their meaning in this context (e.g. "In the last 24 hours, to what extent did you feel like you had the available time to be with your children?”). Responses to these questions were also made using a visual analog scale in the Wear-IT app with anchors “no time” to “plenty of time” and coded on a scale from 0 to 5. Responses from these 6 items were averaged to obtain a parent’s daily parental resources score used in our analyses.

Analysis Plan

Before answering our research questions, we assessed the construct validity of our adapted daily measures by computing the correlations between the baseline pre-EMA measures and the person-level mean values of the equivalent daily measures. We also checked the extent to which our daily measures of parental self-efficacy and children’s home literacy activities exhibited between-person and between-day variability. To do this, we computed summary statistics of the measures and plotted their trajectories across time. We also computed the Intraclass Correlation (ICC; Bolger et al. 2013) to understand the extent to which inter- and intra-individual variation relatively contributed to the observed variability.

We tested our three hypotheses using generalized linear multilevel models that estimated relationships between children’s daily literacy activity levels and parental self-efficacy. Missing data was row-wise omitted from the analysis. Because children’s daily literacy activity levels were counts of time during specific intervals, we employed log-link functions to accommodate their non-normality (Coxe et al., 2009). We ran two models, one predicting levels of parent-child shared reading activities and the other predicting children’s levels of independent reading. Both models had the following form:

$$\log\,(RM_{it}) = \beta_{0i} + \beta_1 PSE_{avg_i} + \beta_2 PSE_{it} + \beta_3 PSE_{i(t-1)} + \{covariates\} + \mu_i + \varepsilon_{it})$$

$$\{covariates\} = \beta_4 CE_i + \beta_5 CS_i + \beta_6 CG_i + \beta_7 TR_{avg_i} + \beta_8 TR_i + \beta_9 RM_{i(t-1)}$$

In the first model, the outcome RM_{it} was the number of minutes parents reported they read with their child that day, in the second model, RM_{it} was the number of minutes parents reported their child read independently that day. PSE_{avg_i} was the person-level mean of parents' daily reports of parental self-efficacy over the 14-day measurement burst. PSE_{it} was the person-centered level of parental self-efficacy that day, and $PSE_{i(t-1)}$ was the person-centered level of parental self-efficacy the previous day. Finally, μ_i was a person-level random intercept and ε_{it} was normally distributed, zero-mean error term.

We included covariates in both models informed by the Hoover-Dempsey & Sandler (1995, 1997) model of parental involvement. At the person-level, we controlled for parent's initial reports of their child's levels of reading enjoyment and struggle (CE_i and CS_i, respectively), as well as their grade level (CG_i). We also controlled for the person-level mean of parents' daily reports of their time resources (TR_{avg_i}) along with their person-centered report of time resources that day. Finally, we controlled for the number of minutes children engaged in the targeted activity the previous day ($RM_{i(t-1)}$).

Results

We recruited 16 parents and they all consented to being in the study. Out of the 16 parents, all identified as female except 1 parent, who identified as male. Parents were an average age of 44.5 (SD = 7) and predominantly identified as white (88%). Most of the children identified as male, except 5 who identified as female. The mean grade children were about to enter was 3.25 (SD = 1.06). Over the course of the 14-day measurement burst, parents submitted a mean of 10.75 (SD = 3.87) surveys. Descriptive statistics for the measures collected at the initial survey along with the means from the daily surveys are shown in Table 2-1. Between-parent correlations of the initial survey and daily survey means are shown in Table 2-2.

Initial measures of parental self-efficacy and child's independent reading were significantly correlated to their daily counterparts ($p < 0.05$), but levels of parental time resources and child shared reading were not ($r_{Time\ Resources}$ = 0.37, p = 0.15; $r_{Shared\ Reading}$= 0.21, p = 0.43). ICCs of all the daily measures were high (0.66 - 0.80), justifying the use of multilevel models in the later analyses. Time series plots of variable trajectories, shown in Figure 2-1, also confirmed high levels of nesting of daily observations within parents. Within-parent correlations of the parent-level-centered daily measures are presented in Table 2-3.

Table 2-1: Descriptive statistics of results from the initial and daily surveys

	Mean	St. Dev.	Min	Max	ICC
Initial Survey					
Parent Age	44.50	7.00	29.00	62.00	
Child grade level	3.25	1.06	2.00	5.00	
Child reading enjoyment	3.36	1.35	1.30	5.00	
Child reading struggle	3.51	1.41	1.00	5.00	
Parental Self-efficacy	5.40	0.82	4.11	7.00	
Parental time resources	3.76	0.92	2.23	5.00	
Shared Reading Frequency	2.69	0.48	2.00	3.00	
Independent Reading Frequency	2.19	0.91	1.00	3.00	
Daily Survey					
Parental Self-efficacy (mean)	3.76	0.85	2.19	4.94	0.80
Parental Time Resources (mean)	2.93	0.77	2.09	4.87	0.66
Shared Reading Minutes (mean)	55.67	47.86	2.14	178.57	0.72
Independent Reading Minutes (mean)	69.44	66.92	0.00	217.14	0.73
N Daily Reports	10.75	3.87	4.00	14.00	

Table 2-2: Between-parent correlations.

	1.	2.	3.	4.	5.	6.	7.	8.	9.
Child reading enjoyment (1)									
Child reading struggle (2)	-0.75*								
Parental Self-efficacy (3)	0.00	0.09							
Parental time resources (4)	0.01	-0.23	0.41						
Shared Reading Frequency (5)	-0.06	0.42	-0.22	-0.35					
Independent Reading Frequency (6)	0.86*	-0.59*	-0.05	-0.08	0.14				
Parental Self-efficacy (mean) (7)	-0.21	0.07	0.51*	0.43	-0.03	-0.31			
Parental Time Resources (mean) (8)	0.18	-0.32	0.04	0.37	-0.14	0.09	0.60*		
Shared Reading Minutes (mean) (9)	0.46	-0.11	0.54*	0.10	0.21	0.53*	0.36	0.28	
Independent Reading Minutes (mean) (10)	0.73*	-0.28	0.25	-0.12	0.12	0.53*	0.00	0.01	0.51*

(*p < 0.05)

Table 2-3: Within-parent correlations (parent-centered)

	1.	2.	3.	4.	5.	6.
Parental Self-efficacy (1)						
Parental Time Resources (2)	0.35*					
Shared Reading Minutes (3)	-0.04	-0.13				
Independent Reading Minutes (4)	0.25*	0.05	0.09			
Parental Self-efficacy (previous day) (5)	-0.14	-0.16*	0.03	0.09		
Shared Reading Minutes (previous day) (6)	0.03	0.02	0.20*	0.01	-0.04	
Independent Reading Minutes (previous day) (7)	-0.12	0.00	0.15	0.14	0.25*	0.09

(*p < 0.05)

Table 2-4: Results from the multilevel models predicting daily levels of children's shared reading and independent reading activities.

	Shared Reading	**Independent Reading**
(Intercept)	-1.82 (1.88)	-5.96 (3.09)
Parental Self-efficacy (mean)	0.89 (0.28)*	0.20 (0.39)
Parental Self-efficacy (parent-centered)	0.01 (0.15)	0.52 (0.10)*
Parental Self-efficacy (parent-centered, prev. day)	-0.09 (0.14)	0.19 (0.09)*
Covariates		
Child reading enjoyment	0.70 (0.19)*	0.90 (0.27)*
Child reading struggle	0.19 (0.20)	0.38 (0.32)
Child grade level	-0.48 (0.21)*	0.57 (0.34)
Parental Time Resources (mean)	-0.67 (0.37)	-0.03 (0.55)
Parental Time Resources (parent-centered)	-0.29 (0.09)*	-0.14 (0.09)
Shared Reading Minutes (prev. day)	0.04 (0.01)*	
Independent Reading Minutes (prev. day)		0.02 (0.01)*
AIC	681.33	681.64
BIC	713.76	714.16
Log Likelihood	-329.66	-329.82
Num. obs.	141	142
Num. groups	16	16
Group σ^2 (Intercept)	0.29	0.67

*p < 0.05

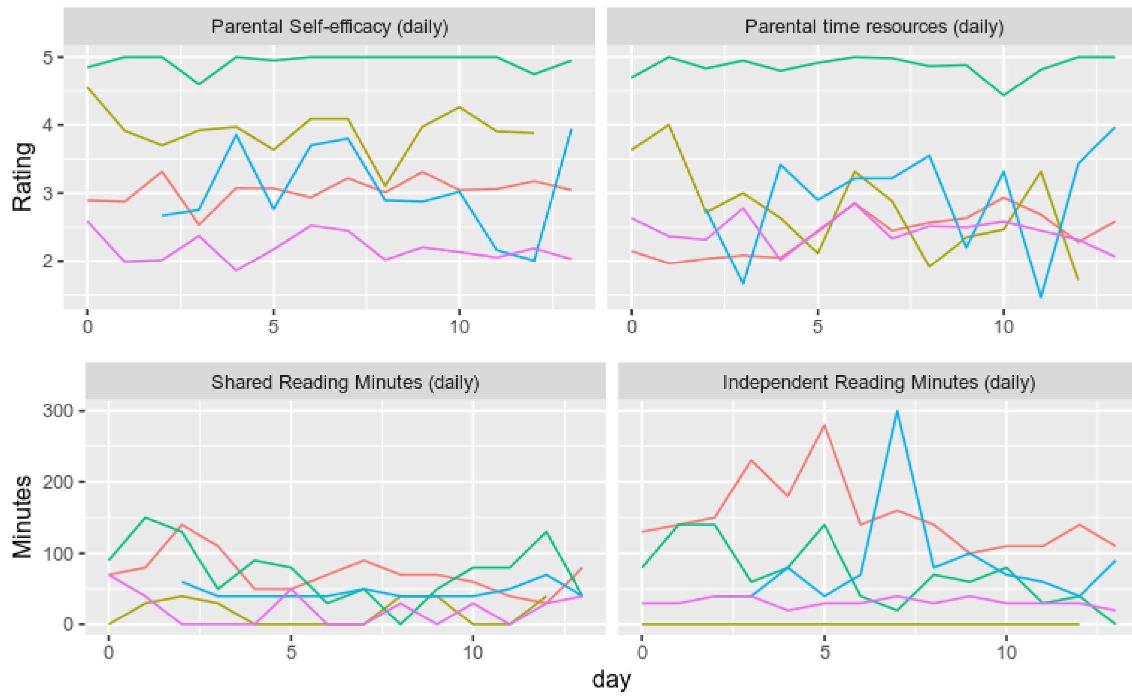

Figure 2-1: Time series plots showing trajectories of five parents

Results from the multilevel models predicting daily levels of children's shared reading and independent reading activities are shown in Table 2-4. For shared reading activities, only a between-parent effect of parental self-efficacy was observed (β_2 = 0.89, SE = 0.28, p < 0.05); no evidence of within-parent effects of same-day or previous-day parental self-efficacy was found. Independent reading activities had the opposite pattern of results, with no evidence of a between-person effect of parental self-efficacy, but significant effects of same-day and previous-day levels of parental self-efficacy (β_3 = 0.52, SE = 0.10, p < 0.05; β_4 = 0.19, SE = 0.09, p < 0.05). Significant covariates of shared reading activities included child levels of reading enjoyment (β4 = 0.70, SE = 0.19, p < 0.05), child grade level (β_6 = -0.48, SE = 0.21, p < 0.05), and levels of shared reading the previous day (β_9 = 0.04, SE = 0.01, p < 0.05). Significant covariates of independent reading activities included child levels of reading enjoyment (β_4 = 0.90, SE = 0.27, p < 0.05), and levels of independent reading the previous day (β_9 = 0.02, SE = 0.01, p < 0.05).

Discussion

Parents play an active role in their child's education by facilitating interactions with their child and motivating their child to engage in activities that support their child's literacy development. Theoretical models of parental involvement suggest that individual differences in levels of involvement are related to individual and contextual processes that unfold across time in the home environment. As process-level theories, they make predictions about the systematic relationships we should expect to observe between parents' daily experiences and levels of involvement with their child's literacy activities. In this study we used an Ecological Momentary Assessment (EMA; Carson et al., 2010) measurement design to obtain daily measures of parents' experience of parental self-efficacy and their children's literacy activities for 16 parents over a 14-day measurement burst. These data enabled us to test theoretical predictions about process-level relationships between parental self-efficacy and children's home literacy activities.

Correlations between the initial assessment and parent-level means of the daily measures supported the construct validity of the daily measures. Results from the PSAM significantly correlated with parent-level means of our two-item day-to-day version ($r = 0.51$, $p < 0.05$), and parents' initial report of their child's typical independent reading frequency significantly reported with the mean of their daily reports ($r = 0.53$, $p < 0.05$). Our 16-parent sample's limited power to detect between-person effects, so it is not surprising correlations between initial reports of parental time resources and child shared reading the means of their daily counterparts did not reach levels of significance ($r = 0.37$, $p = 0.15$; $r = 0.21$, $p = 0.43$). Regardless, the magnitude of their point estimates indicated a general level of convergence between the scales on par with other EMA studies (e.g., Targum et al., 2021).

Our daily measures of parental self-efficacy, time resources, and children's literacy activities exhibited both between- and within-parent variability in the 16 parents across the 14-

day measurement burst. All daily measures showed high intra-class correlations (ICC; Bolger et al. 2013), ranging from 0.66-0.80, indicating that between-parent differences accounted for most of their observed variability, but not all. This was visually confirmed by time series plots of parents' trajectories across the 14-day measurement burst.

In traditional analyses, ICC is often used as an indicator of "test-retest" reliability, where fluctuations in repeated measures of the same individual are interpreted as random variability around a "true" score for that individual (Aldridge et al., 2017). If the within-person variability we observed in our measures was random, then no correlations should exist in the day-to-day fluctuations in the measures. Instead, we found significant correlations in the day-to-day fluctuations in our measures, supporting their interpretation as indicators of processes unfolding time, rather than random errors of measures representing static, parent-level characteristics.

To test our hypotheses about the process-level relationships between parents' parental self-efficacy and children's levels of literacy activities, we built two statistical models. The first model predicted children's daily levels of shared reading activities from daily levels of parental self-efficacy, the second model predicted children's daily levels of independent reading from daily levels of parental self-efficacy. Both models controlled for potential dyad-level and day-level confounders, such as children's enjoyment of reading and a parents' available time resources that day. Our models provided mixed support of our hypotheses.

Our first hypothesis predicted a positive between-person relationship in parental self-efficacy and children's home literacy activities, but this relationship was only significant in the model predicting shared reading activities. Although it was surprising to not find evidence that children of parents with higher levels of parental self-efficacy had higher levels of independent reading activities, it is possible our null result could be explained by the lack of power in our 16-parent sample to detect between-person effects. The theorized mechanism by which we expected higher levels of parental self-efficacy to relate to higher levels of children's independent reading relied on parents influencing their child's behavior. Shared reading, by contrast, is an activity that parents can more directly facilitate and initiate. Therefore, it's very possible a between-person relationship in parental self-efficacy and children's independent reading activities exists in our data but is much smaller than its relationship to shared reading activities and below our threshold to detect. The point estimates from our models tell a similar story: although the estimate of the between-person effect of parental self-efficacy on children's independent reading is positive, it is over 4 times smaller than the between-person effect of parental self-efficacy on children's independent reading.

Our second and third hypotheses predicted within-person relationships in daily levels of parental self-efficacy and children's home literacy activities. On this front, we found that parents' same-day and previous-day's levels of parental self-efficacy significantly predicted their child's daily levels of independent reading activities, but no within-person effects of parental self-efficacy were found for shared reading activities. In this case, insufficient power is not a likely explanation of our null result for shared reading activities, because our power to detect within-person effects was much higher than our power to detect between-person effects. While the between-person estimates of effect could only leverage variability across 16 individuals, the within-person estimates considered variability across the entire pool of days collected by those 16 individuals, for a total of 140 observations. Our results therefore support the hypothesis that parental self-efficacy and children's independent reading activities are related by way of day-to-day processes but suggest that parental self-efficacy and children's shared reading activities do not have similar process-level connections at the day-to-day level.

In our third hypothesis, social cognitive theory and resource allocation theory made divergent predictions. Social cognitive theory predicted a positive relationship between parents' levels of self-efficacy and their child's literacy activities the next day, where resource allocation theory predicted a negative relationship. Our findings suggest that when parents experience success in motivating their children to engage in independent reading activities on a particular day, their success supports their future confidence and ability to motivate their child, rather than decreasing their future attention and performance on the task. In other words, our results supported social cognitive theory's predictions for children's independent reading activities, but contradicted resource allocation theory.

Limitations and Future Directions

As mentioned earlier, our 16-parent sample had limited power to detect between-person relationships, so our between-person findings should not be considered strong evidence for or against our first hypothesis. A replication of this study with a larger sample size would be necessary to shed more light on the nature of our observed between-person relationships in parental-efficacy and children's home literacy activities.

Although our power to detect within-person relationships was substantially greater than our power to detect between-person effects, our relatively small sample of parents limits the generalizability of the within-person results. The parents in our sample were predominately white, well-educated, and were selected by convenience by way of their participation in a summer goal-setting program, rather than by random sampling. It is unclear whether our results would generalize beyond the present sample to a more general population of parents. Future studies need to obtain more diverse samples of parents to study the possibility that parental self-efficacy and children's home literacy activities have different dynamics in other populations of parents.

In addition to our within-person findings having limited generalizability to other populations, they also have limited temporal generalizability. During the summertime, parents and children have different routines and schedules at home as compared to the rest of the school year. It's possible we would have obtained different results had we run the study at a different part of the school year or even during a different part of the summer. Future work should consider multiple burst study EMA designs (Cho et al., 2019) that sample different parts of a school year in order to determine the extent to which our findings can be generalized across seasons or are unique to the processes that drive parent-child interactions around summer reading.

Finally, our results may not generalize across other levels of analysis of time (i.e. time scales). In the present study we only considered dynamics of parental self-efficacy and children's home literacy activities at a day-to-day level. It is not guaranteed that we would find similar patterns of results had we obtained measures at other time scales (e.g., month-to-month or year-to-year; Chen et al., 2005). Current process models of parental involvement do not explicitly consider how these processes may exhibit different dynamics when considered on different time scales, so our choice of a day-to-day time scale was made based on previous work in the study of other emotional processes (e.g., Wallace et al., 2017). More theoretical development that incorporates multiple levels of analysis of temporal dynamics into models of parental involvement is necessary to guide future empirical work in this area (Deboeck et al., 2015).

Conclusion

Although previous work has examined relationships between parental self-efficacy and children's home literacy activities, to our knowledge this is the first study to observe these processes as they unfold at a day-to-day level. Repeated observations at the day-level made it possible to test process-level theories about how individuals' experiences shape their decisions and behavior across time. Our results supported a social cognitive view of the relationships between parental self-efficacy and their children's independent reading activities. We found that when parents experienced success getting their child to read independently, it was related to increases in their self-efficacy that day, as well as predictive of their future success motivating their child's independent reading the next day. In other words, our work supports the view that parental self-efficacy and children's independent reading activities are self-reinforcing processes: that the parents' experience of success motivating their child supports future successes. In education, this is sometimes referred to as a "Matthew effect" (Pfost et al., 2014).

Our work adds to previous studies relating differences in parental characteristics to differences in home literacy activities by considering how those differences may arise as the result of parents' changing experiences across time shaping their behavior. Without considering dynamics across time with repeated observations, it is not possible to distinguish between individual variability related to intrinsic individual characteristics, and individual variability arising from interacting individual and contextual processes as they unfold across time. The present study provides a process-level window into the relationship between parental self-efficacy and children's independent reading activities, explicitly considering how their dynamics relate across time within families.

An understanding of these relationships between parental self-efficacy and children's literacy activities has the potential to help us develop more contextually sensitive family literacy interventions (Baker, 2013; Anderson et al., 2010; van Steensel et al., 2012; Mol et al., 2008). A key challenge for home literacy interventions that target parent involvement as a mechanism of action is often regular adherence from parents (de la Rie et al., 2017; Sawyer et al., 2021). Understanding of how daily fluctuations of parental self-efficacy relate to parents' level of involvement with their children's literacy activities can potentially be used in the design of adaptive home literacy interventions (e.g., Just In Time Adaptive Interventions; Nahum-Shani et al., 2018) to detect when parents could benefit from additional motivational components to support their long-term success and engagement in the intervention.

This work represents a step towards a more detailed, process-level understanding of how parents' daily experiences in the home environment relate to their engagement in their child's literacy development. Although this study focused on relationships in parental self-efficacy and their children's literacy activities, there are a multitude of reasons why parents may find it challenging to engage their child in literacy activities in their home environment, most of which have not yet been studied on a day-to-day level (Sawyer et al., 2021; Weigel et al., 2006; Johnson et al., 2008). Future work is needed to develop and test theories about the dynamics of these processes to give us a more rounded understanding of children's home literacy environments. The better we can understand the nature of the day-to-day challenges (and opportunities!) parents experience in supporting their children's literacy development, the more effective and inclusive we can design home literacy interventions to be.

References

Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing Test-Retest Reliability of Psychological Measures. *European Psychologist*, *22*(4), 207–218. https://doi.org/10.1027/1016-9040/a000298

Anderson, J., Anderson, A., Friedrich, N., & Kim, J. E. (2010). Taking stock of family literacy: Some contemporary perspectives. *Journal of Early Childhood Literacy*, *10*(1), 33–53. https://doi.org/10.1177/1468798409357387

Anderson, K. J., & Minke, K. M. (2007). Parent Involvement in Education: Toward an Understanding of Parents' Decision Making. *The Journal of Educational Research*, *100*(5), 311–323. https://doi.org/10.3200/JOER.100.5.311-323

Baker, C. E. (2013). Fathers' and Mothers' Home Literacy Involvement and Children's Cognitive and Social Emotional Development: Implications for Family Literacy Programs. *Applied Developmental Science*, *17*(4), 184–197. https://doi.org/10.1080/10888691.2013.836034

Ballard, T., Yeo, G., Loft, S., Vancouver, J. B., & Neal, A. (2016). An integrative formal model of motivation and decision making: The MGPM*. *Journal of Applied Psychology*, *101*(9), 1240–1265. https://doi.org/10.1037/apl0000121

Bandura, A. (1977). *Self-efficacy: Toward a Unifying Theory of Behavioral Change*.

Bandura, A. (2000). Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Beck, J. W., & Schmidt, A. M. (2012). Taken out of context? Cross-level effects of between-person self-efficacy and difficulty on the within-person relationship of self-efficacy with resource allocation and performance. *Organizational Behavior and Human Decision Processes*, *119*(2), 195–208. https://doi.org/10.1016/j.obhdp.2012.06.009

Bingham, G. E. (2007). Maternal Literacy Beliefs and the Quality of Mother-Child Book-reading Interactions: Assoications with Children's Early Literacy Development. *Early Education & Development*, *18*(1), 23–49. https://doi.org/10.1080/10409280701274428

Boerma, I. E., Mol, S. E., & Jolles, J. (2018). Parents Adjust the Quality of Their Home Literacy Environment to the Reading Interest of Their Third to Sixth Graders. *Parenting*, *18*(4), 243–261. https://doi.org/10.1080/15295192.2018.1524243

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research* (Illustrated edition). The Guilford Press.

Brick, T. R., Mundie, J., Weaver, J., Fraleigh, R., & Oravecz, Z. (2020). Low-Burden Mobile Monitoring, Intervention, and Real-Time Analysis Using the Wear-IT Framework: Example and Usability Study. *JMIR Formative Research*, *4*(6), e16072. https://doi.org/10.2196/16072

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of Child Psychology*.

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, *37*(4), 408–426. https://doi.org/10.1598/RRQ.37.4.4

Burris, P. W., Phillips, B. M., & Lonigan, C. J. (2019). Examining the Relations of the Home Literacy Environments of Families of Low SES with Children's Early Literacy Skills. *Journal of Education for Students Placed at Risk*, *24*(2), 154–173. https://doi.org/10.1080/10824669.2019.1602473

Carson, R. L., Weiss, H. M., & Templin, T. J. (2010). Ecological momentary assessment: A research method for studying the daily lives of teachers. *International Journal of Research & Method in Education*, *33*(2), 165–182. https://doi.org/10.1080/1743727X.2010.484548

Castro, D. C., Bryant, D. M., Peisner-Feinberg, E. S., & Skinner, M. L. (2004). Parent involvement in Head Start programs: The role of parent, teacher and classroom characteristics. *Early Childhood Research Quarterly*, *19*(3), 413–430. https://doi.org/10.1016/j.ecresq.2004.07.005

Chang, R., & Little, T. D. (2018). Innovations for Evaluation Research: Multiform Protocols, Visual Analog Scaling, and the Retrospective Pretest–Posttest Design. *Evaluation & the Health Professions*, *41*(2), 246–269. https://doi.org/10.1177/0163278718759396

Chen, G., Bliese, P. D., & Mathieu, J. E. (2005). Conceptual Framework and Statistical Procedures for Delineating and Testing Multilevel Theories of Homology. *Organizational Research Methods*, *8*(4), 375–409. https://doi.org/10.1177/1094428105280056

Chen, G., Gully, S. M., Whiteman, J.-A., & Kilcullen, R. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology*, *85*, 835–847. https://doi.org/10.1037/0021-9010.85.6.835

Cho, G., Pasquini, G., & Scott, S. B. (2019). Measurement burst designs in lifespan developmental research. In *Oxford Research Encyclopedia of Psychology*.

Coleman, P. K., & Karraker, K. H. (1998). Self-Efficacy and Parenting Quality: Findings and Future Applications. *Developmental Review*, *18*(1), 47–85. https://doi.org/10.1006/drev.1997.0448

Coxe, S., West, S. G., & Aiken, L. S. (2009). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

de la Rie, S., van Steensel, R. C. M., & van Gelderen, A. J. S. (2017). Implementation quality of family literacy programmes: A review of literature. *Review of Education*, *5*(1), 91–118. https://doi.org/10.1002/rev3.3081

Deboeck, P. R., Nicholson, J., Kouros, C., Little, T. D., & Garber, J. (2015). Integrating Developmental Theory and Methodology: Using Derivatives to Articulate Change Theories, Models, and Inferences. *Applied Developmental Science*, *19*(4), 217–231. https://doi.org/10.1080/10888691.2015.1021924

Dulay, K. M., Cheung, S. K., & McBride, C. (2018). Environmental correlates of early language and literacy in low- to middle-income Filipino families. *Contemporary Educational Psychology*, *53*, 45–56. https://doi.org/10.1016/j.cedpsych.2018.02.002

Dumka, L. E., Stoerzinger, H. D., Jackson, K. M., & Roosa, M. W. (1996). Examination of the Cross-Cultural and Cross-Language Equivalence of the Parenting Self-Agency Measure. *Family Relations*, *45*(2), 216–222. https://doi.org/10.2307/585293

Giallo, R., Treyvaud, K., Cooklin, A., & Wade, C. (2013). Mothers' and fathers' involvement in home activities with their children: Psychosocial factors and the role of parental self-efficacy. *Early Child Development and Care*, *183*(3–4), 343–359. https://doi.org/10.1080/03004430.2012.711587

Green, C. L., Walker, J. M. T., Hoover-Dempsey, K. V., & Sandler, H. M. (2007). Parents' motivations for involvement in children's education: An empirical test of a theoretical model of parental involvement. *Journal of Educational Psychology*, *99*(3), 532–544. https://doi.org/10.1037/0022-0663.99.3.532

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental Involvement in Children's Education: Why Does it Make a Difference? *Teachers College Record*, *97*(2), 310–331. https://doi.org/10.1177/016146819509700202

Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of Educational Research*, *67*(1), 3–42.

Johnson, A. D., Martin, A., Brooks-Gunn, J., & Petrill, S. A. (2008). Order in the House! Associations among Household Chaos, the Home Literacy Environment, Maternal Reading Ability, and Children's Early Reading. *Merrill-Palmer Quarterly (Wayne State University. Press)*, *54*(4), 445–472. https://doi.org/10.1353/mpq.0.0009

Jones, T. L., & Prinz, R. J. (2005). Potential roles of parental self-efficacy in parent and child adjustment: A review. *Clinical Psychology Review*, *25*(3), 341–363. https://doi.org/10.1016/j.cpr.2004.12.004

Kahneman, D. (1973). *Attention and Effort*.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657–690. https://doi.org/10.1037/0021-9010.74.4.657

Luo, R., Pace, A., Levine, D., Iglesias, A., de Villiers, J., Golinkoff, R. M., Wilson, M. S., & Hirsh-Pasek, K. (2021). Home literacy environment and existing knowledge mediate the link between socioeconomic status and language learning skills in dual language learners. *Early Childhood Research Quarterly*, *55*, 1–14. https://doi.org/10.1016/j.ecresq.2020.10.007

Lynch, J., Anderson, J., Anderson, A., & Shapiro, J. (2007). Parents' Beliefs About Young Children's Literacy Development And Parents' Literacy Behaviors. *Reading Psychology*. http://www.tandfonline.com/doi/abs/10.1080/02702710500468708

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added Value of Dialogic Parent–Child Book Readings: A Meta-Analysis. *Early Education and Development*, *19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Nesselroade, J. R., & Ram, N. (2004). *Studying Intraindividual Variability: What We Have Learned That Will Help Us Understand Lives in Context*. 21.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, *9*(3–4), 427–440. https://doi.org/10.1016/0885-2006(94)90018-3

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual Differences in Reading Development: A Review of 25 Years of Empirical Research on Matthew Effects in Reading. *Review of Educational Research*, *84*(2), 203–244. https://doi.org/10.3102/0034654313509492

Raikes, H., Alexander Pan, B., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., Banks Tarullo, L., Abigail Raikes, H., & Rodriguez, E. T. (2006). Mother?Child Bookreading in Low-Income Families: Correlates and Outcomes During the First Three Years of Life. *Child Development*, *77*(4), 924–953. https://doi.org/10.1111/j.1467-8624.2006.00911.x

Rodriguez, E. T., Tamis-LeMonda, C. S., Spellmann, M. E., Pan, B. A., Raikes, H., Lugo-Gil, J., & Luze, G. (2009). The formative role of home literacy experiences across the first three years of life in children from low-income families. *Journal of Applied Developmental Psychology*, *30*(6), 677–694. https://doi.org/10.1016/j.appdev.2009.01.003

Rowe, M. L., Pan, B. A., & Ayoub, C. (2005). Predictors of Variation in Maternal Talk to Children: A Longitudinal Study of Low-Income Families. *Parenting*, *5*(3), 259–283. https://doi.org/10.1207/s15327922par0503_3

Sawyer, B. E., Cycyk, L. M., Sandilos, L. E., & Hammer, C. S. (2018). 'So many books they don't even all fit on the bookshelf': An examination of low-income mothers' home literacy practices, beliefs and influencing factors. *Journal of Early Childhood Literacy*, *18*(3), 338–372. https://doi.org/10.1177/1468798416667542

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, *60*, 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Schwarzer, R., & Luszczynska, A. (2008). *Self efficacy*.

Sénéchal, M., & LeFevre, J.-A. (2014). Continuity and Change in the Home Literacy Environment as Predictors of Growth in Vocabulary and Reading. *Child Development*, *85*(4), 1552–1568. https://doi.org/10.1111/cdev.12222

Sénéchal, M., Lefevre, J.-A., Hudson, E., & Lawson, E. (1996). Knowledge of Picture-Books as a Predictor of Young Children's Vocabulary. *Journal of Educational Psychology*, *88*, 520–536. https://doi.org/10.1037/0022-0663.88.3.520

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Silinskas, G., Sénéchal, M., Torppa, M., & Lerkkanen, M.-K. (2020). Home Literacy Activities and Children's Reading Skills, Independent Reading, and Interest in Literacy Activities From Kindergarten to Grade 2. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01508

Singh, L., Yeung, W.-J. J., Cheng, Q., & Heng, E. Y.-T. (2022). The home literacy environment mediates effects of socio-economic status on infant vocabulary development. *Developmental Science*, *n/a*(n/a), e13349. https://doi.org/10.1111/desc.13349

Targum, S. D., Sauder, C., Evans, M., Saber, J. N., & Harvey, P. D. (2021). Ecological momentary assessment as a measurement tool in depression trials. *Journal of Psychiatric Research*, *136*, 256–264. https://doi.org/10.1016/j.jpsychires.2021.02.012

Tazouti, Y., & Jarlégan, A. (2019). The mediating effects of parental self-efficacy and parental involvement on the link between family socioeconomic status and children's academic achievement. *Journal of Family Studies*, *25*(3), 250–266. https://doi.org/10.1080/13229400.2016.1241185

Van Horn, M. L., Bellis, J. M., & Snyder, S. W. (2001). Family Resource Scale-Revised: Psychometrics and Validation of a Measure of Family Resources in a Sample of Low-Income Families. *Journal of Psychoeducational Assessment*, *19*(1), 54–68. https://doi.org/10.1177/073428290101900104

Van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading*, *29*(4), 367–382. https://doi.org/10.1111/j.1467-9817.2006.00301.x

van Steensel, R., Herppich, S., McElvany, N., & Kurvers, J. (2014). How Effective Are Family Literacy Programs for Children's Literacy Skills? In *Handbook of Family Literacy*. Routledge. https://doi.org/10.4324/9780203841495.ch9

Vancouver, J., Thompson, C., & Williams, A. (2001). The Changing Signs in the Relationships among Self-efficacy, Personal Goals, and Performance. *The Journal of Applied Psychology*, *86*, 605–620. https://doi.org/10.1037/0021-9010.86.4.605

Walker, J. M. T., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental Involvement: Model Revision through Scale Development. *The Elementary School Journal*, *106*(2), 85–104. https://doi.org/10.1086/499193

Wallace, M. L., McMakin, D. L., Tan, P. Z., Rosen, D., Forbes, E. E., Ladouceur, C. D., Ryan, N. D., Siegle, G. J., Dahl, R. E., Kendall, P. C., Mannarino, A., & Silk, J. S. (2017). The role of day-to-day emotions, sleep, and social interactions in pediatric anxiety treatment. *Behaviour Research and Therapy*, *90*, 87–95. https://doi.org/10.1016/j.brat.2016.12.012

Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83. https://doi.org/10.1037/met0000030

Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Mothers' literacy beliefs: Connections with the home literacy environment and pre-school children's literacy development. *Journal of Early Childhood Literacy*, *6*(2), 191–211. https://doi.org/10.1177/1468798406066444

Weinberger, J. (1996). *Literacy Goes to School: The Parents' Role in Young Children's Literacy Learning*. SAGE.

Wu, C., Gong, X., Luo, L., Zhao, Q., Hu, S., Mou, Y., & Jing, B. (2021). Applying Control-Value Theory and Unified Theory of Acceptance and Use of Technology to Explore Pre-service Teachers' Academic Emotions and Learning Satisfaction. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.738959

Yeo, G. B., & Neal, A. (2006). An examination of the dynamic relationship between self-efficacy and performance across levels of analysis and levels of specificity. *Journal of Applied Psychology*, *91*(5), 1088–1101. https://doi.org/10.1037/0021-9010.91.5.1088

Yeo, G. B., & Neal, A. (2013). Revisiting the Functional Properties of Self-Efficacy: A Dynamic Perspective. *Journal of Management*, *39*(6), 1385–1396. https://doi.org/10.1177/0149206313490027

Yeo, L. S., Ong, W. W., & Ng, C. M. (2014). The Home Literacy Environment and Preschool Children's Reading Skills and Interest. *Early Education and Development*, *25*(6), 791–814. https://doi.org/10.1080/10409289.2014.862147

Chapter 3
Using Technology to Enable Low-burden, Intensive Longitudinal Measurement in Family Literacy Programs

Family literacy programs (FLPs) can be versatile way to provide additional support to children who are not responding sufficiently to classroom instruction or at risk for developing reading disabilities (McConnell et al., 2016; van Steensel et al., 2011; Erion, 2006; Gortmaker et al., 2007; Casey et al., 2011). FLPs represent a diverse range of intervention approaches, ranging from training parents in home-based literacy activities with their child, to home-school partnerships (e.g. Sénéchal and Young, 2008). Here, we focus on the former kind of program, in which adult family members are trained in simple, evidence-based reading activities (e.g. phonics activities, repeated reading, etc.) they can incorporate into their daily life with their child. Partnering with parents to support their children's literacy development in their home environment is advantageous because students can receive support that they may otherwise not receive at school due to scheduling, resource, and personnel limitations of providing individualized attention to children directly (Hollands et al., 2013). Furthermore, implementations of FLPs are flexible and customizable, as evidenced by their success in adapting to a wide range of socioeconomic and cultural contexts (Fikrat-Wevers et al., 2022; Murad & Topping 2000; Knauer et al., 2020).

Despite the advantages of FLPs, enactment of parent-child literacy activities can be challenging for parents (Justice et al., 2015; de la Rie et al., 2017). There are a multitude of reasons why parents may find it difficult to regularly engage their child in literacy activities in their home environment, including conflict with their child, how busy they are, and their beliefs about their ability to make an impact on their child's educational outcomes (de Jong et al., 2022;

Sawyer et al., 2018; Weigel et al., 2006; Johnson et al., 2008). At the end of the day, even if the cognitive mechanisms driving the effect of an FLP work miracles, its real-world impact and scalability will be limited if the mechanisms driving FLP adoption and integration are not also well understood (Fixsen et al., 2005). As is often argued in implementation science, studying the factors and processes affecting FLP implementation is critical to their real-world impact, sustainability, and scalability (Durlak & DuPre, 2008).

Parents' and children's affective experiences as they engage in the literacy activities in the FLP are two such individual processes that may affect enactment across time during an FLP. Process models of parent involvement with their children's education based in social cognitive theory (e.g. Hoover-Dempsey & Sandler, 1997; Green et al., 2007; Walker et al., 2005; Anderson & Minke, 2007) suggest parents' experiences of success enacting the activities of an FLP with their child will produce positive feelings that day and will translate into increased levels of future engagement with the intervention (Schunk & DiBenedetto, 2020; Bandura, 1977; Bandura, 2000). Similarly, their children's negative experiences during enactment sessions may modulate parents' future levels of enactment by way of children resisting enactment sessions or creating parent-child conflicts that discourage parental motivation (de Jong et al., 2022). Tests of these theories as they apply to the individual dynamics of FLP enactment require the capability to simultaneously collect intensive longitudinal measures of FLP enactment and parent-child affective experiences during FLP activity sessions.

Although obtaining contemporaneous measures of literacy activity enactment and parent and child affective experiences during an FLP is possible with traditional measurement designs (e.g. parent logs or video and audio recordings), it involves significant parental burden, research cost and introduces potential sources of measurement bias. In the present study we test the feasibility of instead using technology-based materials in an FLP to automatically collect these measures during implementation, without the usual trade-offs of data, burden, and bias common in traditional approaches. Although technology has a long history of use in literacy intervention, previous work has largely focused on the dissemination and delivery of intervention activities (e.g. Barratt-Pugh et al., 2022; DuBois, et al., 2014; Jamshidifarsani et al., 2019). By contrast, the use of mobile technology to support FLP implementation measurement has not received similar attention. This is surprising given the present ubiquity of mobile technology and its ability to automatically collect large amounts of data with low cost and burden (Neef, 2014).

Using the data from this pilot, we tested predictions made by process models of parental involvement about within-person relationships between parents' and children's emotional experiences and literacy activity enactment. We present the results of these analyses and discuss implications for the study of these processes involved in FLP implementations and discuss how this measurement approach could be used to study other potentially salient individual and contextual processes. Finally, we discuss how future work might approach using mobile technology for measurement and evaluation in other FLPs.

Background

Variability in FLP implementation is commonly conceptualized as implementation fidelity (Powell & Cary, 2012; de la Rie et al., 2017). Implementation fidelity (also commonly termed implementation "quality" or "integrity") is broadly defined as the extent to which a program is implemented according to its design (Carroll et al., 2007). Powell and Cary's (2012) conceptual framework of FLP implementation fidelity defines three key elements of FLP implementation: delivery, receipt, and enactment. From these elements, measures of FLP implementation fidelity are derived based on the element's quantity and quality.

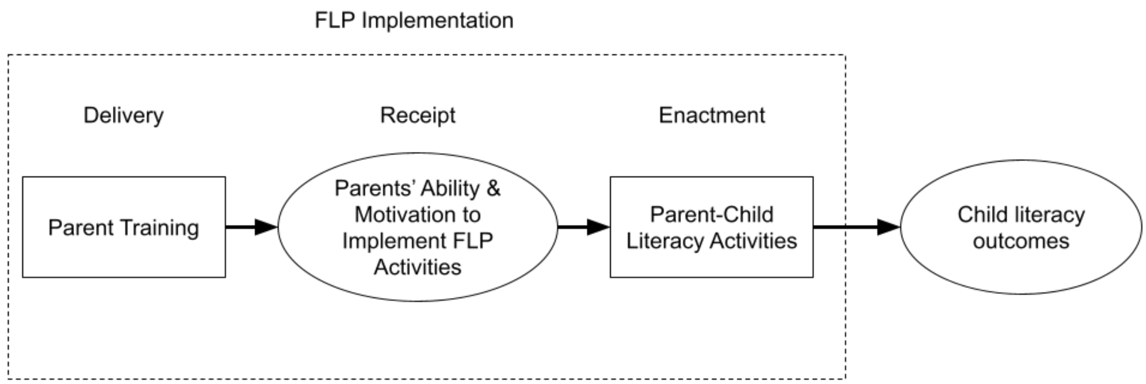

Figure 3-1: A general conceptual model for FLPs, highlighting elements of implementation as defined by Powell and Cary (2012). Boxes represent intervention activities, circles represent outcomes of activities.

Table 3-1: Examples measures of FLP delivery, receipt, and enactment “quantity” and “quality”

	Quantity	**Quality**
Delivery	Number & duration of parent training sessions	Proportion of training steps followed by intervention staff; responsiveness to parents’ questions
Receipt	Number of training sessions parents attend	Parents levels of attention during training; parents’ ability to enacting the literacy activities with their child
Enactment	Number & duration of parent-child activity sessions	Parent adherence to the literacy activities’ procedures; parents’ responsiveness to the needs of their child

As illustrated in Figure 3-1, delivery, receipt, and enactment reflect sequential steps along a causal chain connecting FLP implementation activities to child outcomes. FLPs do not target children directly; instead they more proximally target parents by way of parent training. Therefore, the first element of FLP implementation fidelity the delivery of training to parents. The proximal effects of parent training include increases in parent ability and motivation to implement literacy activities with their child, which defines parents’ receipt of the FLP. These effects of parent training then enable parents to regularly enact what they learned by regularly engaging their child in literacy activities in the home environment, which are designed to proximally target children's literacy outcomes. Each element along the FLP implementation causal chain is necessary for the FLP to impact child literacy outcomes. As shown in Table 3-1, measures of implementation fidelity are derived by considering the quantity and quality of each element involved in FLP implementation.

In the evaluation of FLPs, measures of parent enactment are arguably the most important measures of implementation to capture. As measures of the final step in the implementation causal chain, they can act as a global indicator for the success of the FLP implementation; the key interface between an FLP implementation and the child-level outcomes being targeted. The parent-child activities in an FLP are the proximal causes of the FLP's effect on child outcomes, representing the "core engine" of the intervention's theory of change. If an implementation of an FLP is not having its intended effect on child outcomes, the parent-child activities in an FLP are the first place to look for implementation issues. Measures of enactment make it possible to distinguish between failures of intervention implementation and failures of the FLP's theory of change. To illustrate this, consider an extension of the intervention "engine" metaphor: if your car isn't moving, the first thing you might want to check is your engine. If your engine is not turning, you can deduce that the reason you are stuck is because of a failure in the causal chain between the ignition and the engine starting (i.e. the FLP's implementation), or a failure of the engine's turning to drive the wheels (i.e. the FLP's mechanism for affecting children's literacy trajectories).

A Process View of Parent Enactment

By definition, parents' enactment of FLP literacy activities with their child is itself a process, characterized by meaningful change and variation as the result of bidirectional interactions between the parent, their child, and their environment (Bronfenbrenner & Morris, 2006). Empirical work supports a process view of FLP enactment, with evidence of extensive implementation variability in FLPs existing between and within families across time. Justice et al. (2015), for example, examined reports from caregiver reading log submissions during their 30-week FLP in a latent class analysis, and found four distinct patterns of parental engagement: "completers", parents who maintained high levels of intensity throughout the program; "nonstarters", parents who dropped out at the beginning of the program; "late dropout", parents who dropped out of the program about three-quarters of the way through; and "sporadic", parents who exhibited high levels of engagement at the beginning and end of the intervention, but were minimally involved in the middle. While many FLPs may be designed (and evaluated) with the hope that most parents will fall into the "completer" category, Justice et al. found that parents of this description represented only about a third of their total sample of parents.

Justice et al.'s work underscores the need to study real-world implementations of FLPs with an understanding of parent enactment as a complex, dynamic process in itself, not simply as a static moderator of an FLP's effects on children's literacy development. As a process, parent enactment will be affected by other individual and contextual processes simultaneously unfolding in the home environment during an FLP. An understanding of these other individual and contextual processes that support or derail parent enactment is key to our ability to design FLPs that show robust effects across implementations (Joyce & Cartwright, 2020).

Process models of parent involvement provide insight into the individual processes that may support or derail parent enactment during an FLP. The Hoover-Dempsey & Sandler (1995, 1997) model of parental involvement in education and its later iterations (Green et al., 2007; Walker et al., 2005; Anderson & Minke et al., 2007) describe two determinants of parental involvement that may relate to parents' and children's experiences during enactment sessions: parents' motivational beliefs and parents' perceptions of invitations for involvement.

In the Hoover-Dempsey & Sandler model of parent involvement, parents' motivational beliefs are understood through the lens of social cognitive theory (Bandura, 2000). Social cognitive theory suggests that parents' self-efficacy in FLP enactment, broadly defined as their perception of their skill or ability in enactment, will be increased by way of positive, "mastery experiences" of enactment during the FLP (Schwarzer & Luszczynska 2008). Social cognitive theory predicts that these positive experiences will translate into future levels of motivation to engage their children in enactment activities. This suggests that when parents have positive experiences during FLP enactment it will increase their motivation to initiate enactment sessions in the future with their child.

Similarly, children's experiences during literacy activities in an FLP may also play a role in parents' enactment across time. In the Hoover-Dempsey & Sandler model of parent involvement, children can affect parents' level of involvement in their education by positively or negatively encouraging their parents to engage in involvement activities. When children have negative experiences during the literacy activities of an FLP, this may affect future parent enactment by way of parents expecting children to resist enactment activities in the future,

Unfortunately, it is difficult to obtain reliable measures of FLP enactment and parents' and children's experiences during enactment sessions to study these processes as they unfold across time. Enactment occurs inside the home environment, and its schedule is different for each family as dictated by the rhythms of parents' daily lives, schedules, and the general chaos of family life. In practice, obtaining measures of enactment in FLPs has traditionally involved considerable tradeoffs in the amount of data collected, bias, and burden.

Limitations Of Traditional Measurement Approaches

A common approach for obtaining measures of parent enactment in FLPs is to ask parents to keep written diaries, checklists, or logs of their literacy activity sessions with their child (e.g. Resetar et al., 2006; Gortmaker et al., 2007; Casey et al., 2011; Kupzyk et al., 2012; Lonigan & Whitehurst 1998). This, unfortunately, increases parental burden by giving parents additional tasks to complete, manage, track, and submit during their activity sessions with their child. Given that parents frequently cite reasons related to “time pressures” as one of the key barriers to FLP implementation (Justice et al., 2015), it is likely that measurement designs relying on increases in parental burden may also have an impact on the motivational and contextual processes we wish to observe, resulting in non-random missingness.

Increases in parent burden may also result in selection biases by way of parent attrition, causing parents with less time or motivation to drop out of the study. Parent attrition is already an issue for many FLPs: in a review of the implementation quality of FLPs, de la Rie et al. (2017) found that reports of parent attrition in FLP evaluations varied considerably, some reaching as high as 60%. Even when parents don’t drop out, parents may simply omit tasks that create extra burden: Lonigan & Whitehurst (1998), for example, reported that 40% of the parents in their FLP did not return their reading logs to the intervention staff. Measuring parent implementation of FLPs by way of parent report creates a direct trade-off between the amount of implementation data we collect, and the amount of potential bias introduced in our data (Brick, 2020).

Another measurement approach in the study of FLP enactment is to have parents make video or audio recordings of their activity sessions with their child (Resetar et al., 2006; Gortmaker et al., 2007; Casey et al., 2011; Kupzyk et al., 2012). This approach has the advantage of reducing the amount of additional “paperwork” required by parents to document their activities, and allowing researchers to directly observe parent-child interactions as they occur in context. On the other hand, setting up the video or audio recording equipment is still an extra, “only for research” step parents must remember during implementation, and is highly intrusive from a privacy standpoint.

The use of measures obtained from video and audio recordings also involves considerable research burden in the form of increased costs, time, and potential for data entry errors. In order to obtain and use video or audio recordings in an evaluation of an FLP, researchers must allocate additional money in their research budget to purchase audio / video equipment, securely transfer and store the audio / video data after it is collected, and then hire additional research staff to code the media files. These costs present a significant barrier to the use of video and audio recordings in large-scale evaluations of FLPs. In practice, it is not uncommon for evaluations of FLPs that collect video or audio recordings of parent activities with their child to study parent implementation to have small sample sizes or only code a small percentage of the recordings (e.g. 25% inResetar et al., 2006; 34% in Gortmaker et al., 2007).

Finally, direct, in-person observation of parent-child literacy activities is also possible. Researchers can observe parents enacting the FLP literacy activities with their child in a school, clinic, or lab environment, or travel to their home and observe how they enact literacy activities in their home environment (e.g. Weitzman et al., 2004). Out of all the approaches described so far, this is the most invasive and resource intensive. Although it enables researchers to directly observe the quality of parent-child activities during literacy activities, the researcher’s presence reduces the ecological validity of these observations. Furthermore, this approach is less useful for understanding processes related to frequencies and quantities of parent enactment over the course of the intervention, because only a small number of parent-child literacy activity sessions are observed.

Measurement By Digital Exhaust

A limitation shared by all the aforementioned FLP measurement approaches is that they all require adding additional steps to intervention protocols for the purpose of generating research data (e.g., keeping and submitting reading logs, starting and stopping recording devices, etc.). These extra procedural steps rarely provide value to parents or children aside from the incentives researchers provide for participating in their study. When parents get busy or are already feeling burdened, these "extra" steps are the easiest to omit or forget.

A novel measurement paradigm developed by Klinkenberg et al., (2011) suggests that implementation and measurement in an FLP need not require separate procedural steps, but can occur simultaneously, transparent to the parents and children engaging with intervention. In an approach they called Computer Adaptive Practice (CAP), Klinkenberg et al. (2011) used data automatically obtained by their online e-learning platform "Math Garden" to estimate cognitive ability in real time and adapt intervention items according to students' individual needs. Klinkenberg et al.'s platform achieved this by having their software keep track of every item delivered to students as they engaged with the e-learning platform, and the students' responses. These passively obtained measures of individual performance were used in an item-response model to simultaneously estimate students' mathematics ability and the difficulty of the items in real time (Hofman et al., 2018).

The success of Math Garden highlights technology's potential to combine implementation and measurement during an intervention. In general, interventions making use of technology can program their software for the dual purpose of facilitating the intervention's implementation while simultaneously keeping detailed logs of individuals' interactions with the technology during intervention implementation. Data collected in this manner, sometimes called "digital exhaust" (Neef, 2014), requires no additional participant burden to obtain, and can be automatically sent to researchers over the internet for analysis. Although Klinkenberg et al.'s (2011) work focused on the use of these data for studying cognitive processes, their platform also collected other forms of digital exhaust such as the dates and times when students used the intervention, the length of time they spent in each activity and students' response latency for each item they responded to during intervention activity sessions. These data provided a wealth of opportunities to study other individual processes as they interacted with the Math Garden intervention, such as problem skipping behaviors (Savi et al., 2018; 2021), students' activity preferences (Brinkhuis et al., 2020), and program attrition (Broeke et al., 2022).

Measurement By Ecological Momentary Assessment

In addition to measures derived from digital exhaust, technology can facilitate low-burden measurement by way of ecological momentary assessment (EMA; Shiffman et al., 2008; Carson et al., 2010). EMA is a measurement approach by which short surveys are delivered to individuals during the course of their regular activities and routines in life by way of mobile phones or other ubiquitous mobile devices. The EMA approach offers extreme flexibility in measurement designs, enabling the study of a wide variety of processes across context and time. For example, a "two-week measurement burst" design might deliver a daily survey to a sample of individuals in order to observe daily fluctuations across a particular set of variables across time, or an EMA study might use sensor information on an individual's mobile device to detect their geographic location and deliver surveys relevant to that location (e.g. Depp et al., 2019). Obtaining responses from individuals about their experiences in context minimizes the recall biases found in traditional retrospective surveys (Solhan et al., 2009). EMA is a well-established method in the behavioral sciences, used to study individual processes ranging from the dynamics of emotion (Kuppens et al., 2010) to physical activity (Dunton et al., 2017).

EMA designs provide another way by which we can combine measurement and intervention using technology. Just-in-time adaptive interventions (JITAI), for example, is an intervention approach that uses mobile technology to attempt to deliver the right prompts and the right time to help change behavior (Nahum-Shani et al., 2018). These interventions take advantage of digital exhaust that is continuously being collected by individuals' mobile devices and use it to determine optimal moments of intervention based on decision rules (Klasnja et al., 2015). This approach has shown success in targeting outcomes including physical activity (Hardeman et al., 2019) and addictive behaviors (Goldstein et al., 2017).

As a proven way to study individuals in their ecological contexts, EMA measurement approaches are an attractive way to study the individual and contextual processes affecting enactment activities of FLPs. In the present study, we pilot this measurement technique to obtain low-burden measures of parents' and children's emotional experiences during literacy activity sessions.

Technology in FLP Literacy Activities

Intervention measurement via the digital exhaust and EMA approaches are the most "transparent", that is, least burdensome, when they are integrated into the intervention activities in some way. In Math Garden, for example, technology delivered the entire intervention, and so the measurement via digital exhaust was completely invisible to students. For FLPs, however, replacing parent-child enactment activities with technology-delivered activities children can complete individually is less practical, because their theories of change often include social and cultural components (Bloome, 1985). When parents engage their children in literacy activities, they are not solely engaging in processes affecting their cognitive development, but also modeling other literacy-related behaviors which are important aspects of long-term reading development (e.g., an interest in reading and ways of engaging with a text; Hume et al., 2015). Also, in-person, social delivery by parents has the additional advantage of enabling parents to flexibly adapt their enactment of literacy activities to meet different needs, values, and cultural expectations (Anderson et al., 2010). Therefore, technology that completely replaced parents' enactment of literacy activities with their child would undermine these important intervention supports and ancillary targets of FLPs. That said, technology does not need to replace social interactions in parent enactment in order to be utilized in an FLP: instead, it can be integrated into the intervention's materials.

Previous work has demonstrated technology's potential for being integrated into the materials used by the parents enacting literacy activities found in FLPs, in ways that do not compromise social interactions. In fact, in some cases these technology-enhanced materials have been shown to increase parent-child interactions (Troseth et al., 2020). Parent-child literacy activities found in FLPs typically fall into two categories: coding-focused activities, which target children's decoding and word recognition skills, and comprehension-focused activities, which target children's language development and critical thinking skills (van Steensel et al., 2011). Although technology has been successfully integrated into the materials used in both types of activities, its potential for extending our measurement capabilities has not been explicitly studied.

Volpe et al.'s (2009, 2011) "Tutoring Buddy" software is an example of how technology-enhanced materials can be used in code-focused literacy activities found in FLPs. Volpe et al.'s intervention used software designed to assist tutors as they engaged children in letter-sound recognition activities in an incremental rehearsal practice procedure (Varma & Schleisman 2014). Volpe et al.'s technology did not replace tutor-child interactions; it instead replaced the flashcards and tile manipulatives traditionally used in such interventions. Unlike traditional flashcards or tile manipulatives, however, Volpe et al.'s software had the ability to scaffold tutor-child intervention procedures and track children's individual responses across time in the intervention. Although

Volpe et al. did not describe using other forms of digital exhaust to test process-level theories about implementation, Tutoring Buddy indicates the possibility for FLP implementation measurement by way of digital exhaust in coding-focused FLP activities.

Work by Troseth et al. (2020) demonstrates how technology can be integrated into comprehension-focused activities in an FLP. Troseth et al. had parents use e-readers in shared reading activities with their children that automatically scaffolded dialogic reading lines of questioning designed to target comprehension and language outcomes. Troseth et al.'s intervention featured a character named Ramone from the PBS Kids Peg + Cat television program who would pop up on the page while parents and children were engaged in reading activities and model elements from dialogic questioning (Towson et al., 2017) about the story. Although Troseth et al.'s intervention demonstrates how technology can be integrated into shared reading activities in an FLP, they unfortunately did not use the full extent of digital exhaust that the e-reader could potentially generate, because this capability wasn't included in their software implementation. Instead, they adopted the more traditional measurement approach of manually coding video and audio recordings of parent-child interactions.

Present Study

In the present study, we tested the feasibility and usability of using technology-based intervention materials to study the relationship between parents' enactment of an FLP across time and parents' and children's emotional experiences during the enactment activities. To do this, we developed a proof-of-concept mobile app to replace the paper materials of an existing code-focused FLP, the Syllabee parent training program (Stevenson, 2018). Our custom mobile app passively collected digital exhaust measures of parents' implementation and child performance during literacy activities along with EMA measures of parents' and children's emotional experiences. We piloted the app with three parents in the Syllabee parent training program, who used the app for approximately half of a school year. We present pilot data automatically collected by the app during the pilot, parent ratings of the app's acceptability, and selected quotes by parents about their experiences using the app. With these data we tested the following hypotheses:

Hypothesis 1: Parents will find the mobile app acceptable and usable, as defined by a user experience rating above 7 on a 10-point scale. Also, qualitative results will confirm that our measurement approach did not interfere with the enactment activities or discourage parents' use of the app.

Hypothesis 2: Parents' and children's experiences of positive emotions during the enactment of literacy activities with their child will be associated with fewer days elapsing before their next day of intervention activities with their child.

After testing these hypotheses, we conclude by discussing the implications of our results, and describe how variations of our technology-enhanced FLP measurement approach could be used to study other enactment-related processes and in other implementations of FLPs. Finally, we consider how technology supporting the integration of intervention implementation and measurement may play a role in the development of more contextually sensitive FLPs.

Method

Participants

Parent-child dyads from the Portland, OR metropolitan area were recruited through the Syllabee parent training program (Stevenson, 2018), an existing parent tutoring program in reading that trains parents in word-level literacy activities to strengthen their child's word recognition ability. Parents provided their informed consent and children gave their assent to use their anonymized data collected by the app for research purposes.

Materials

The Syllabee parent training program is a parent tutoring intervention designed for children struggling in word-level reading ability. Parents are trained in simple activities to target word-level recognition skills and are encouraged to regularly practice them with their children. The activities in the intervention focused on developing decoding skills through direct and systematic phonics instruction, which has shown to help children struggling with reading (Ehri et al., 2001).

Through our partnership with Syllabee, we developed a mobile app designed to replace the paper worksheets used in one of the program's main activities. In the activity's original form, parents used the provided worksheet to prompt the child with nonwords to practice decoding, and were trained in error handling procedures designed to encourage self-correction. A progression of worksheets at increasing levels of difficulty allowed parents to systematically introduce new phonics rules into the practice as the child gained mastery. These worksheets were also used in a variant of this activity, where parents would dictate nonsense words and children would practice spelling them. Our mobile app provided the same functionality as the worksheets while passively capturing measures obtained via digital exhaust from the app, including the duration, content, and item-level response data from the literacy activity sessions.

To begin the activity, parents selected the desired level of difficulty for their practice session, and the app would begin generating nonsense words at that difficulty or below. Parents then could interact with the app as they would interact with a deck of flash cards. The device would display a single word at a time, allowing the parent to prompt the child to decode the word

(Figure 3-2). Upon child response, parents would swipe left to get a new word, or swipe up to initiate an error-handling procedure guided by the app. It was up to parents to decide how much practice to complete with their children based on their time and energy levels. When the parent indicated they were finished with the activity, the app delivered a brief survey to record the parent and child's affect and fatigue levels (Figure 3-3).

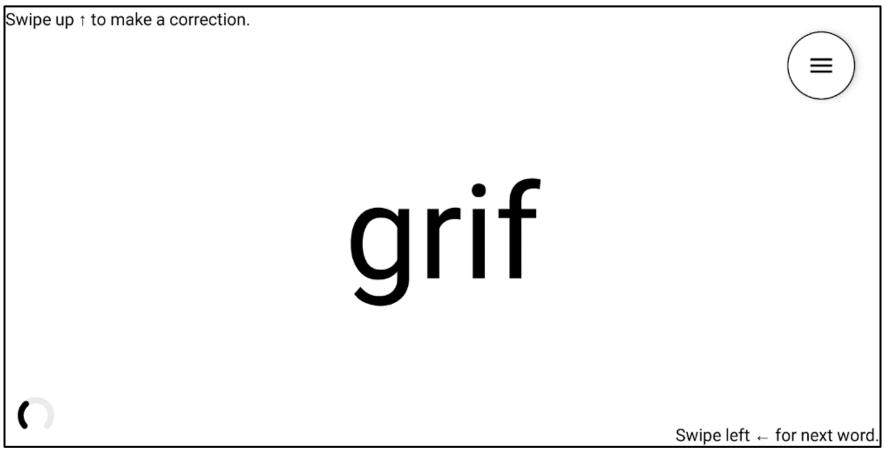

Figure 3-2: An example screen from the app showing a nonsense word

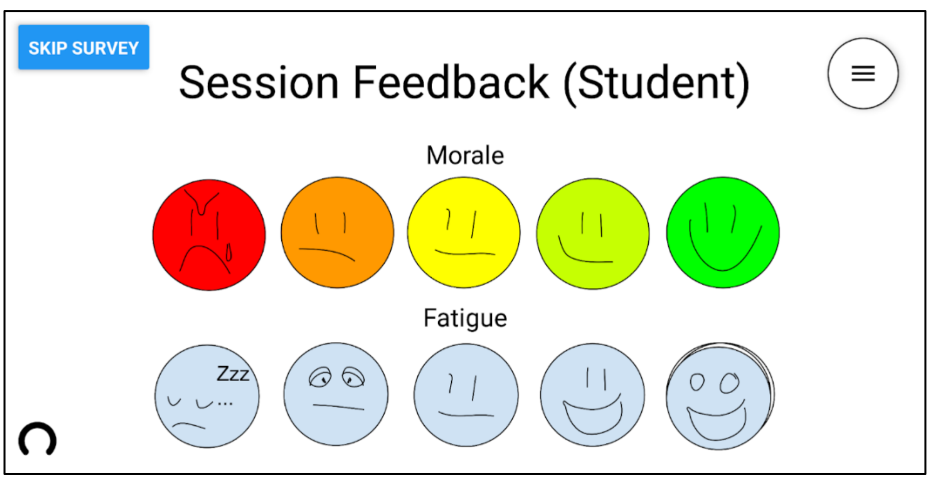

Figure 3-3: Self-reported parent and child morale and fatigue screen.

Procedures

Three parent-child dyads in the Syllabee parent training program volunteered to beta-test our app for their tutoring sessions. The app automatically logged their session times and durations, along with the words displayed, corrections made, and brief surveys of morale and fatigue levels after activities. During weekly check-ins with their reading specialist in the Syllabee training program, parents were given the opportunity to share any feedback about their use of the app and encouraged to report any bugs or difficulties they had using it. The reading specialist took notes during the parent responses and relayed them to the research team. Although the content of the training program only lasted eight weeks, parents continued meetings with the Syllabee reading specialist for up to 15 weeks. After parents graduated from the Syllabee program, they were encouraged to continue using the app for the rest of the school year.

Measures

Parent Acceptability: During their weekly sessions with their Syllabee reading specialist, parents were asked to rate their experience using the app on a likert scale from 1 (very frustrating) to 10 (enjoyable and easy to use).

Activity Timing and Duration: The app automatically recorded the time and date when activities were started and stopped. (The start time was used in analysis.)

Word Difficulty Level: Nonsense words for activity sessions were automatically generated by the app by drawing randomly from grapheme pools of word beginnings, middles and endings and filtering out unpronounceable or inappropriate word constructions. Word difficulty during sessions could thus be modulated by restricting or adding to these pools of available graphemes. Parents would select a desired difficulty level prior to starting an activity session with their child according to a preset, 20-level difficulty progression developed by the Syllabee parent training program.

Item-level Response: Our app recorded all items delivered during tutoring sessions. When children did not recognize words correctly, parents initiated an app-facilitated error handling procedure. This allowed the app to passively record a binary indicator for each item indicating if the child correctly recognized the word. In our analyses, we divided the number of correct responses by the total number of items delivered in literacy activities that day, to compute the "correct item response rate".

Affective Experience During Literacy Activities: Parents and children self-reported morale and fatigue levels after each activity session on 1-5 likert scales using the screen shown in Figure 3-3. Parents were instructed to let their children provide their own self-ratings, rather than assessing the ratings for them. We used only the reports of morale in our analyses.

Data Analysis Plan

To test the acceptability of the mobile app, we reviewed parent reports of their weekly experiences with the app, recorded during their weekly meetings with the reading specialist during the 8-week program. We computed the mean rating of their experience using the app, for each parent and used a single-sided t-test to test if it was greater than a 7 out of 10. We also examined notes and quotes recorded weekly by the reading specialist regarding parents' experience with the app.

We then computed summary statistics and made time-series plots of the intensive longitudinal measures automatically collected by the app. We plotted both the raw data, as well as the data with a month-wide smoothing window to get a sense of the fast- and slow-time structure of variability captured by the app over the course of the FLP.

Finally, we used the data collected by the app to test our hypotheses about relationships between adherence and parents' and children's affective experiences during the literacy activity sessions in the FLP. We tested both hypotheses using generalized linear models predicting the number of days between subsequent parent-child literacy activity sessions from their reports of morale during activities. To test hypothesis 2, we used two models, one estimating the effect of parent morale, and the other estimating the effect of child morale. Reports of morale were mean-centered at the person-level, to isolate the within-person effect.

Both models used log-link functions to account for the non-normal distribution of the count data (Coxe et al., 2009). In addition to mean person-level morale levels, we included covariates of the number of days in the program, the duration of the activities that day, the activity's level of difficulty, and the percent of items answered correctly. When parents and children recorded multiple sessions of literacy activities in the same day, we aggregated their sessions by summing the activity durations and number of items and taking the mean of their reports of morale across sessions that day and the activity difficulty level. We also dropped the row with the largest number of days recorded between literacy activities to account for possible outliers introduced if parents took a break from literacy activities with their child during winter holidays.

Results

Three parents from the Syllabee parent training program volunteered to beta-test the app. Their demographic information is presented in Table 3-2. All three parents were mothers; two were high school graduates, and one had a college degree. The children ranged from 9-10 years old; one child was male, and the other two were female.

On average, parents rated their experiences with the app a 9.3 out of 10 (SD = 0.8), across 13.0 (SD = 3.5) weekly meetings with the Syllabee reading specialist. A single-sided t-test indicated the average rating across the three parents was significantly greater than a 7 out of 10 ($t = 5.0$, $df = 2$, $p = 0.02$). The app was enthusiastically received by parents, who reported that the app made it more convenient to integrate the literacy activities into their daily routines, and "represented a missing piece" in their child's education. One of the parents, who had previously used the paper-based materials, reported that the app was "SO much simpler than using the sheets of paper". Another parent noted that the app would fit well in a public-school setting. To understand why parents rated the app's usability on some weeks higher than others, we reviewed the notes recorded by the Syllabee reading specialist. These notes suggested that parents' lower ratings of the app were only related to early bugs in the software that were fixed during development, and not due to any burdens caused by our data collection approach.

Parents reported using the technology-based materials in unique ways that enhanced parent-child interactions during literacy activities in ways we did not expect. One parent, for example, reported that their child enjoyed holding the mobile phone during literacy activities and swiping the words, which they used as a way to motivate their child during the literacy activities. Perhaps most unexpectedly, parents found that the post-activity surveys were a useful tool to check in with their child during activity sessions. Parents reported these check-ins helped avoid conflicts arising from frustration and fatigue. One parent reported that when their child indicated that their energy was low, they were able to amicably work together to end the activity session when they reached their limit, without "melting down". This novel use of the post-activity surveys was shared by the Syllabee reading specialist with the other parents. The other parents reported similar success employing this strategy with their children.

Usage data automatically collected by the app confirmed parents' reports that they used the app regularly and incorporated the intervention activities into their weekly routines. Figure

3-4 shows time series plots of app usage across time for the three dyads, and Table 3-3 shows descriptive statistics of measures automatically collected by the dyad's interactions with the mobile app. The three participating dyads used the app for a total of 25-28 weeks, with total app usage times ranging from 4-10 hours. Dyad app usage abruptly dropped in March 2020 with the start of the global COVID-19 pandemic, although a few dyads still did a few tutoring sessions with the app as late as April 2020. Figure 3-5 shows time series plots of morale and fatigue ratings recorded the three dyads.

Results from the models predicting the number of days between activities are presented in Table 3-4. Both day-level parent morale and child morale had a significant, negative relationship with the number of days between literacy activity sessions ($\beta_{\text{parent morale}}$ = -0.29, SE = 0.08, $p < 0.05$; $\beta_{\text{child morale}}$ = -0.19, SE = 0.05, $p < 0.05$).

Table 3-2: Demographic information from the three parent-child dyads

Dyad	Parent Gender	Parent Ethnicity	Parent Education	Child Age	Child Grade	Child Gender
Dyad 1	Female	Decline to answer	High school grad	10	Fourth	Female
Dyad 2	Female	White	College grad	10	Fifth	Male
Dyad 3	Female	White	High school grad	9	Fourth	Female

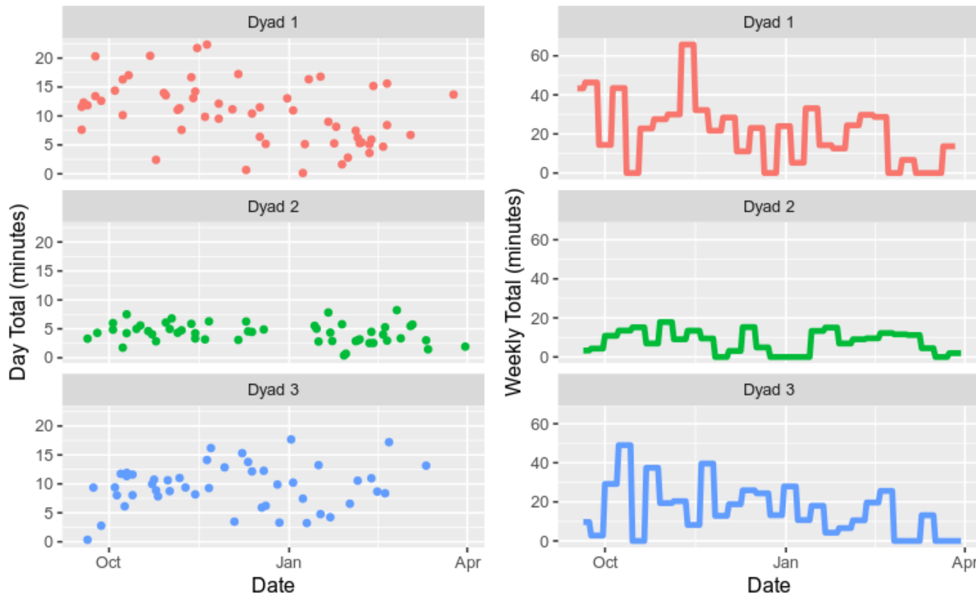

Figure 3-4: Plots of daily app usage across time. The left column plots the total amount of time parents spent engaging their children in literacy activities each day they used the app. The right column aggregates the daily reports to show week-level totals.

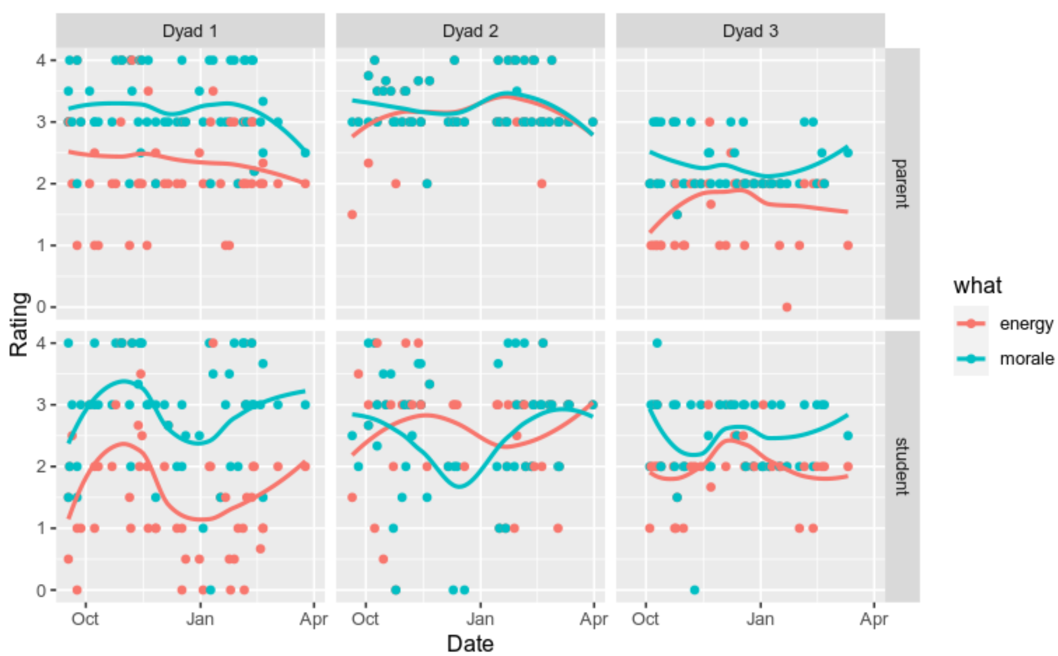

Figure 3-5: Time series plots of morale and fatigue ratings for parents and their children recorded across time for the three dyads, both raw data (points) and smoothed data (loess) to show general trends.

Table 3-3: Descriptive results of measures collected by the app

	Dyad 1	Dyad 2	Dyad 3
App Engagement			
Total number of weeks	27	28	25
Total days with activities	57	52	47
Total minutes of activities	603	223	447
Mean days between activities (sd)	3.39 (3.74)	3.78 (4.70)	3.76 (3.39)
Item-Level Responses			
Total number of items delivered	1735	2410	1285
Item-level correct response rate	85%	95%	89%
Post Activity Survey			
Mean child energy (sd)	1.68 (1.12)	2.56 (0.91)	1.99 (0.53)
Mean parent energy (sd)	2.38 (0.83)	3.19 (0.58)	1.65 (0.62)
Mean child affect (sd)	2.91 (0.92)	2.56 (1.03)	2.50 (0.66)
Mean parent affect (sd)	3.23 (0.64)	3.28 (0.45)	2.30 (0.45)

Table 3-4: Results from models predicting the number of days between activities

	Model 1	Model 2
Parent Morale (person-centered)	-0.29 (0.08)*	
Child Morale (person-centered)		-0.19 (0.05)*
Covariates		
Parent Morale (person-level mean)	-0.04 (0.11)	
Child Morale (person-level mean)		-0.55 (0.34)
Week in program	0.02 (0.01)*	0.03 (0.01)*
Duration of activities (day-level)	0.02 (0.01)*	0.03 (0.01)*
Activity difficulty (day-level)	-0.00 (0.01)	-0.02 (0.01)
Correct item response rate (day-level)	1.56 (0.52)*	1.76 (0.58)*
AIC	852.28	848.19
BIC	873.16	869.08
Log Likelihood	-419.14	-417.10
Deviance	446.28	442.19
Num. obs.	146	146

*$p < 0.05$

Discussion

Parent-child literacy activities are the theoretical mechanism by which FLPs proximally affect child literacy outcomes. Process models of parent involvement suggest parents' and children's experiences during enactment sessions may affect future enactment sessions across time. In this study we tested the feasibility and usability of using technology-based intervention materials to study the dynamic relationships between parent enactment across time in an FLP and parents' and children's affective experiences during enactment sessions.

We developed a mobile app that replaced paper materials in an existing FLP and piloted it with three parent volunteers. In addition to replacing the paper materials used in the literacy activities, the app passively collected digital exhaust measures of FLP enactment. These measures included the timing and duration of parent-child literacy activities, the content covered in the activities, and item-level response data. Unlike traditional measures, obtaining these measures did not require any additional burden from parents and researchers. In addition to digital exhaust measures of enactment, the app also automatically collected EMA measures of parent and child emotional experiences during activity sessions.

Our results supported our first hypothesis that parents would find the app acceptable and usable. In their weekly meetings with the Syllabee reading specialist, parents reported high levels of satisfaction with the app, and indicated it was easy to use and integrate into their daily routines. Not only were the EMA measures not burdensome to parents, but parents also found them to be useful during enactment sessions. Like Troseth et al.'s (2020) technology-based shared reading intervention, these findings indicate the potential for technology to support and extend positive parent-child interactions, rather than replace them.

The app collected passive, digital exhaust measures of parents' enactment in the FLP across time, without the need for parents to manually keep logs of their parent-child activity sessions. Graphs of these data revealed considerable individual variability in all measures within families across time. These measures were more detailed and intensive than measures typically obtained in traditional collection methods. Instead of relying on parents to estimate or remember to observe enactment times and durations, start and end time of each activity was precisely logged by the app as well as the level of activity difficulty parents selected. These data made it possible to directly observe parents' enactment trajectories across time with high levels of precision and without increasing parent burden or introducing retrospective reporting biases. In addition to collecting data via digital exhaust, the app also delivered short, low-burden EMA surveys to parents during enactment activities to study relationships in parents' and children's emotional experiences during activity sessions and their levels of engagement with the FLP across time. Graphs of these data also revealed considerable individual variability within families across time.

Our second hypothesis, based on process models of parent involvement, predicted that parents' and children's positive emotional experiences during enactment would be associated with fewer days until the next day of enactment. In line with our hypotheses, we found that both parents' and children's positive emotional experiences during enactment sessions were associated with fewer days elapsing until their next enactment session. Although these results are limited in generalizability due to our small pilot sample, they demonstrate the theoretical utility of the novel FLP enactment data collected by our technology-based materials.

Limitations & Future Directions

Our small pilot sample size limits the generalizability of the findings of our analyses, so its theoretical significance is limited. Future work would need to repeat the analyses in the present study with a larger sample in order to produce more generalizable findings. We share these results as a demonstration of the types of theoretical questions about FLP enactment that can be answered using a measurement approach that combines implementation and measurement using technology. Although the present study focused on demonstrating proof of concept, the technology-based materials we developed are easier to deploy in larger FLP implementations than traditional data collection approaches. If we were using parent activity logs to collect this information, for example, every additional participant added to the study would produce a significant amount of data that would have to be coded or transcribed by the research staff. With our technology-based approach, these data are automatically collected and sent to researchers in a form that is ready for analysis.

In the present study, we only considered parents' and children's emotional experiences during enactment sessions. Future work should consider how parents' and children's emotional experiences during the day outside of enactment sessions also affect FLP enactment. Future studies might consider collecting these measures by way of EMA assessments during the day.

We also did not obtain measures of parents' feelings of time pressures, which previous work has shown to be a considerable factor affecting parent enactment (Justice et al., 2015). Fluctuations in parents' experiences of time pressures across time represents a potential unmeasured confounder in our analyses, but also is a salient process worthy of its own study. Future work may additionally consider obtaining these measures by way of EMA both during enactment sessions and on days when no enactment occurs in order to better understand the dynamic role it may play in parents' enactment across time.

More work is also necessary to consider how measurement via technology-based materials in FLP generalizes to other populations of parents. Although growing ubiquity of mobile devices paints a promising picture for the use of mobile technology with a wide range of populations in a wide range of environments, future studies will need to test these approaches on samples of individuals more representative of the wider population.

Finally, in this study we focused on reducing cost and burden in measurement of FLP enactment by using technology-based intervention materials but did not consider the cost and resources required to develop and maintain the technology in the first place. Future work will need to conduct formal cost analyses to characterize the trade-offs between the resources required to develop and maintain custom software for FLP implementations and the amount of savings in research cost and burden are achieved. In the long-term, it may be worthwhile for researchers to combine resources and to invest in developing an e-learning platform for FLPs like Math Garden that has the capability to be customized and extended by researchers to support many types of FLP activities and modes of measurement.

Future Technological Extensions

In addition to adding more forms of EMA to our measurement design to measure other enactment-affecting processes of interest (e.g., parent time pressures), we might consider taking advantage of other forms of data exhaust we could collect using our mobile app. We could use GPS location or accelerometer data, for example, to better understand the role parents' physical location and routines play in their initiation of enactment activities with their children. Previous work in EMA designs has pioneered the use of many other wearable sensor technologies to passively collect information that could be used in the study of FLP enactment (Moshe et al., 2021). Children could wear wristbands (e.g., Mahapatra et al., 2022) that would make it possible to detect when they were in close proximity to their parents by way of the parents' phones. Combined with other EMA measures, these kinds of technologies could give us an unprecedented, ecologically valid view into the daily interactions of parents and their children, and the processes that support or derail their enactment sessions in an FLP.

We might also consider using salient measures of enactment-related processes as tailoring variables in a Just-In-Time Adaptive Intervention to support parent enactment of FLP literacy activities (JITAI; Nahum-Shani et al., 2018). A micro-randomized measurement design (Klasnja et al., 2015) would allow us to empirically determine the optimal conditions (e.g., contexts, times, circumstances, etc.) to give parents reminders to enact an FLP's literacy activities with their child (or deliver other micro-interventions to support enactment). Such designs can simultaneously estimate individual dynamics as well as generalizable group trends.

In addition to considering adaptation between sessions to support parent enactment processes, we might also consider forms of adaptation within enactment sessions. Like Math Garden, our technology-based intervention materials collected item-level responses and item-level response timing that could be used to dynamically estimate item difficulties and children's progress across time (Klinkenberg et al. 2011). Future work might also consider extending computer adaptive practice models to incorporate the bidirectional role parents' and children's affective experiences potentially play on children's performance during enactment sessions into activity and item recommendations.

Item-level response and timing data also make it possible to study the individual-level cognitive processes involved in the development of reading. Some theories of developmental dyslexia suggest the existence of subtypes of dyslexia that are related to item-level error patterns (e.g., Friedmann & Coltheart, 2016). Collecting children's item-level responses during practice activities across time opens the possibility of testing these theories using novel statistical techniques such as cognitive diagnostic models (CDM; Li et al., 2016), and data mining (Koedinger et al., 2015).

In this study we created versions of paper materials using mobile technology, but there are many other ways technology could be integrated into the materials used in FLP activities. For example, Tangible User Interfaces (TUI; Rodić & Granić, 2022) are an emerging area of research in which physical objects are used to interact with technology, in ways that do not require the use of screens. Given that many literacy activities often already use physical manipulatives, it may be possible to create technology-enhanced versions of the exact same physical objects that also have the capacity to scaffold parent-child interactions and automatically obtain implementation measures. Alternatively, the use of existing physical objects might be extended with mobile technology by way of Augmented Reality (AR; Pan et al., 2021). With AR, images are projected onto physical objects allowing users to interact with the physical world as a computer interface.

AR is a natural fit for home literacy activities, given previous work demonstrating its success in informal learning environments (Sommerauer & Müller, 2014).

Conclusion

As is often argued in implementation science, studying the factors and processes affecting FLP implementation is critical to their real-world impact, sustainability, and scalability (Durlak & DuPre 2008). Although obtaining measures of FLP enactment has traditionally involved tradeoffs in the amount of data collected, bias, and burden, technology-based materials have the potential to lower barriers to the study of these implementation and implementation-related processes. In this study, we demonstrated the feasibility of using technology-based materials for obtaining detailed and intensive measures of implementation and implementation-related processes in an FLP. We also demonstrated the utility of the resulting data for testing process-level theories about FLP enactment in example analyses that examined relationships between parents' and children's emotional experiences during literacy activities in an FLP and their enactment behavior. Although our results for this FLP measurement approach are promising, future work will be necessary to determine the extent to which this measurement approach generalizes across a wider population of parents and can be adapted to other activities found in FLPs.

References

Anderson, J., Anderson, A., Friedrich, N., & Kim, J. E. (2010). Taking stock of family literacy: Some contemporary perspectives. *Journal of Early Childhood Literacy*, *10*(1), 33–53. https://doi.org/10.1177/1468798409357387

Anderson, K. J., & Minke, K. M. (2007). Parent Involvement in Education: Toward an Understanding of Parents' Decision Making. *The Journal of Educational Research*, *100*(5), 311–323. https://doi.org/10.3200/JOER.100.5.311-323

Bandura, A. (1977). *Self-efficacy: Toward a Unifying Theory of Behavioral Change*.

Bandura, A. (2000). Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064

Barratt-Pugh, C., Hill, S., Johnson, N. F., Barblett, L., & Parker, A. (2022). Designing and Implementing a Family Literacy Program Through Smartphones: How Does Recruitment Method Influence Uptake and Attrition? *Early Childhood Education Journal*. https://doi.org/10.1007/s10643-022-01433-z

Bloome, D. (1985). Reading as a Social Process. *Language Arts*, *62*(2), 134–142.

Brick, T. R., Mundie, J., Weaver, J., Fraleigh, R., & Oravecz, Z. (2020). Low-Burden Mobile Monitoring, Intervention, and Real-Time Analysis Using the Wear-IT Framework: Example and Usability Study. *JMIR Formative Research*, *4*(6), e16072. https://doi.org/10.2196/16072

Brinkhuis, M., Cordes, W., & Hofman, A. (2020). Governing games Adaptive game selection in the Math Garden. *ITM Web of Conferences*, *33*, 03003. https://doi.org/10.1051/itmconf/20203303003

Broeke, N. ten, Hofman, A. D., Kruis, J., de Mooij, S. M. M., & van der Maas, H. (2022). *Predicting and Reducing Quitting in Online Learning* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/htzvm

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of Child Psychology*.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1), 40. https://doi.org/10.1186/1748-5908-2-40

Carson, R. L., Weiss, H. M., & Templin, T. J. (2010). Ecological momentary assessment: A research method for studying the daily lives of teachers. *International Journal of Research & Method in Education*, *33*(2), 165–182. https://doi.org/10.1080/1743727X.2010.484548

Casey, L. B., & Williamson, R. (2011). Training Parents as Effective Literacy Tutors: Increasing the Procedural Integrity of Tutoring. *Mentoring & Tutoring: Partnership in Learning*, *19*(3), 257–276. https://doi.org/10.1080/13611267.2011.597118

Coxe, S., West, S. G., & Aiken, L. S. (2009). The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

de Jong, P. F., Schreurs, B. G. M., & Zee, M. (2022). Parent–child conflict during homeschooling in times of the COVID-19 pandemic: A key role for mothers' self-efficacy in teaching. *Contemporary Educational Psychology*, *70*, 102083. https://doi.org/10.1016/j.cedpsych.2022.102083

de la Rie, S., van Steensel, R. C. M., & van Gelderen, A. J. S. (2017). Implementation quality of family literacy programmes: A review of literature. *Review of Education*, *5*(1), 91–118. https://doi.org/10.1002/rev3.3081

Depp, C. A., Bashem, J., Moore, R. C., Holden, J. L., Mikhael, T., Swendsen, J., Harvey, P. D., & Granholm, E. L. (2019). GPS mobility as a digital biomarker of negative symptoms in schizophrenia: A case control study. *Npj Digital Medicine*, *2*(1), Article 1. https://doi.org/10.1038/s41746-019-0182-1

DuBois, M. R., Volpe, R. J., & Hemphill, E. M. (2014). A Randomized Trial of a Computer-Assisted Tutoring Program Targeting Letter-Sound Expression. *School Psychology Review*, *43*(2), 210–221. https://doi.org/10.1080/02796015.2014.12087446

Dunton, G. F. (2017). Ecological Momentary Assessment in Physical Activity Research. *Exercise and Sport Sciences Reviews*, *45*(1), 48–54. https://doi.org/10.1249/JES.0000000000000092

Durlak, J. A., & DuPre, E. P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology*, *41*(3), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, *71*(3), 393–447.

Erion, J. (2006). Parent Tutoring: A Meta-Analysis. *Education and Treatment of Children*, *29*(1), 79–106. JSTOR.

Fikrat-Wevers, S., van Steensel, R., & Arends, L. (2021). Effects of Family Literacy Programs on the Emergent Literacy Skills of Children From Low-SES Families: A Meta-Analysis. *Review of Educational Research*, 0034654321998075. https://doi.org/10.3102/0034654321998075

Fixsen, D. L. (2005). *Implementation research: A synthesis of the literature*. National Implementation Research Network.

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistics Perspectives*.

Goldstein, S. P., Evans, B. C., Flack, D., Juarascio, A., Manasse, S., Zhang, F., & Forman, E. M. (2017). Return of the JITAI: Applying a Just-in-Time Adaptive Intervention Framework to the Development of m-Health Solutions for Addictive Behaviors. *International Journal of Behavioral Medicine*, *24*(5), 673–682. https://doi.org/10.1007/s12529-016-9627-y

Gortmaker, V. J., Daly III, E. J., McCurdy, M., Persampieri, M. J., & Hergenrader, M. (2007). Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent-tutoring interventions. *Journal of Applied Behavior Analysis*, *40*(2), 203–221.

Green, C. L., Walker, J. M. T., Hoover-Dempsey, K. V., & Sandler, H. M. (2007). Parents' motivations for involvement in children's education: An empirical test of a theoretical model of parental involvement. *Journal of Educational Psychology*, *99*(3), 532–544. https://doi.org/10.1037/0022-0663.99.3.532

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 31. https://doi.org/10.1186/s12966-019-0792-7

Hofman, A., Jansen, B., de Mooij, S., Stevenson, C., & van der Maas, H. (2018). A Solution to the Measurement Problem in the Idiographic Approach Using Computer Adaptive Practicing. *Journal of Intelligence*, *6*(1), 14. https://doi.org/10.3390/jintelligence6010014

Hollands, F. M., Pan, Y., & Shand, R. (2013). *improving early literacy: Cost-effectiveness analysis of effective reading programs*. 73.

Hoover-Dempsey, K. V., & Sandler, H. M. (1995). Parental Involvement in Children's Education: Why Does it Make a Difference? *Teachers College Record*, *97*(2), 310–331. https://doi.org/10.1177/016146819509700202

Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of Educational Research*, *67*(1), 3–42.

Hume, L. E., Lonigan, C. J., & McQueen, J. D. (2015). Children's literacy interest and its relation to parents' literacy-promoting practices. *Journal of Research in Reading*, *38*(2), 172–193. https://doi.org/10.1111/j.1467-9817.2012.01548.x

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education*, *128*, 427–451.

Johnson, A. D., Martin, A., Brooks-Gunn, J., & Petrill, S. A. (2008). Order in the House! Associations among Household Chaos, the Home Literacy Environment, Maternal Reading Ability, and Children's Early Reading. *Merrill-Palmer Quarterly (Wayne State University. Press)*, *54*(4), 445–472. https://doi.org/10.1353/mpq.0.0009

Joyce, K. E., & Cartwright, N. (2020). Bridging the Gap Between Research and Practice: Predicting What Will Work Locally. *American Educational Research Journal*, *57*(3), 1045–1082. https://doi.org/10.3102/0002831219866687

Justice, L. M., Logan, J. R., & Damschroder, L. (2015). Designing Caregiver-Implemented Shared-Reading Interventions to Overcome Implementation Barriers. *Journal of Speech, Language, and Hearing Research : JSLHR*, *58*(6), S1851–S1863. https://doi.org/10.1044/2015_JSLHR-L-14-0344

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Micro-Randomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *34*(0), 1220–1228. https://doi.org/10.1037/hea0000305

Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, *57*(2), 1813–1824. https://doi.org/10.1016/j.compedu.2011.02.003

Knauer, H. A., Jakiela, P., Ozier, O., Aboud, F., & Fernald, L. C. H. (2020). Enhancing young children's language acquisition through parent–child book-sharing: A randomized trial in rural

Kenya. *Early Childhood Research Quarterly*, *50*, 179–190. https://doi.org/10.1016/j.ecresq.2019.01.002

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *WIREs Cognitive Science*, *6*(4), 333–353. https://doi.org/10.1002/wcs.1350

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. https://doi.org/10.1037/a0020962

Kupzyk, S., Iii, E. J. D., & Andersen, M. N. (2012). Preparing Teachers to Train Parents to Use Evidence-based Strategies for Oral Reading Fluency with Their Children. *Contemporary School Psychology*, *16*, 12.

Li, H., Hunter, C. V., & Lei, P.-W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391–409. https://doi.org/10.1177/0265532215590848

Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly*, *13*(2), 263–290. https://doi.org/10.1016/S0885-2006(99)80038-6

Mahapatra, S., Kannan, V., Seshadri, S., Ravi, V., & Sofana Reka, S. (2022). An IoT-Based Wristband for Automatic People Tracking, Contact Tracing and Geofencing for COVID-19. *Sensors (Basel, Switzerland)*, *22*(24), 9902. https://doi.org/10.3390/s22249902

McConnell, B. M., & Kubina, R. (2016). Parents Using Explicit Reading Instruction with Their Children At-Risk for Reading Difficulties. *Education and Treatment of Children*, *39*(2), 115–139. https://doi.org/10.1353/etc.2016.0010

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.625247

Murad, C. R., & Topping, K. J. (2000). Parents as Reading Tutors for First Graders in Brazil. *School Psychology International*, *21*(2), 152–171. https://doi.org/10.1177/0143034300212003

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Neef, D. (2014). *Digital Exhaust: What Everyone Should Know about Big Data, Digitization and Digitally Driven Innovation*. Pearson Education.

Pan, Z., López, M. F., Li, C., & Liu, M. (2021). Introducing augmented reality in early childhood literacy learning. *Research in Learning Technology*, *29*. https://doi.org/10.25304/rlt.v29.2539

Powell, D. R., & Carey, A. J. (2012). Approaches to program fidelity in family literacy research. In *Handbook of family literacy* (pp. 403–416). Routledge.

Resetar, J. L., Noell, G. H., & Pellegrin, A. L. (2006). Teaching parents to use research-supported systematic strategies to tutor their children in reading. *School Psychology Quarterly*, *21*(3), 241–261. https://doi.org/10.1521/scpq.2006.21.3.241

Rodić, L. D., & Granić, A. (2022). Tangible interfaces in early years' education: A systematic review. *Personal and Ubiquitous Computing*, *26*(1), 39–77. https://doi.org/10.1007/s00779-021-01556-x

Savi, A. O., Ruijs, N. M., Maris, G. K., & van der Maas, H. L. (2018). Delaying access to a problem-skipping option increases effortful practice: Application of an A/B test in large-scale online learning. *Computers & Education*, *119*, 84–94.

Savi, A. O., van Klaveren, C., & Cornelisz, I. (2021). *Combating Effort Avoidance in Computer Adaptive Practicing: Does a Problem Skipping Restriction Promote Learning?*

Sawyer, B. E., Cycyk, L. M., Sandilos, L. E., & Hammer, C. S. (2018). 'So many books they don't even all fit on the bookshelf': An examination of low-income mothers' home literacy practices, beliefs and influencing factors. *Journal of Early Childhood Literacy*, *18*(3), 338–372. https://doi.org/10.1177/1468798416667542

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, *60*, 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Schwarzer, R., & Luszczynska, A. (2008). *Self efficacy*.

Sénéchal, M., & Young, L. (2008). The Effect of Family Literacy Interventions on Children's Acquisition of Reading From Kindergarten to Grade 3: A Meta-Analytic Review. *Review of Educational Research*, *78*(4), 880–907. https://doi.org/10.3102/0034654308320319

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Solhan, M. B., Trull, T. J., Jahng, S., & Wood, P. K. (2009). Clinical Assessment of Affective Instability: Comparing EMA indices, questionnaire reports, and retrospective recall. *Psychological Assessment*, *21*(3), 425–436. https://doi.org/10.1037/a0016869

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education*, *79*, 59–68. https://doi.org/10.1016/j.compedu.2014.07.013

Stevenson, K. (2018). *Syllabee Parent Empowerment Dyslexia Remediation*. syllabee.org

Towson, J. A., Fettig, A., Fleury, V. P., & Abarca, D. L. (2017). Dialogic Reading in Early Childhood Settings: A Summary of the Evidence Base. *Topics in Early Childhood Special Education*, *37*(3), 132–146. https://doi.org/10.1177/0271121417724875

Troseth, G. L., Strouse, G. A., Flores, I., Stuckelman, Z. D., & Russo Johnson, C. (2020). An enhanced eBook facilitates parent–child talk during shared reading by families of low socioeconomic status. *Early Childhood Research Quarterly*, *50*, 45–58. https://doi.org/10.1016/j.ecresq.2019.02.009

van Steensel, R., McElvany, N., Kurvers, J., & Herppich, S. (2011). How Effective Are Family Literacy Programs?: Results of a Meta-Analysis. *Review of Educational Research*, *81*(1), 69–96. https://doi.org/10.3102/0034654310388819

Varma, S., & Schleisman, K. B. (2014). The Cognitive Underpinnings of Incremental Rehearsal. *School Psychology Review*, *43*(2), 222–228.

Volpe, R., Burns, M., Dubois, M., & Zaslofsky, A. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools*, *48*, 332–342. https://doi.org/10.1002/pits.20557

Volpe, R. J. (2009). Tutoring buddy. *Rockport, MA: Twin Lights Ltd*.

Walker, J. M. T., Wilkins, A. S., Dallaire, J. R., Sandler, H. M., & Hoover-Dempsey, K. V. (2005). Parental Involvement: Model Revision through Scale Development. *The Elementary School Journal*, *106*(2), 85–104. https://doi.org/10.1086/499193

Weitzman, C. C., Roy, L., Walls, T., & Tomlin, R. (2004). More Evidence for Reach Out and Read: A Home-Based Study. *Pediatrics*, *113*(5), 1248–1253. https://doi.org/10.1542/peds.113.5.1248

Chapter 4

A Vision for the Future of Measurement in Educational Research

In the education sciences, the typical boundaries of generalizability in our theories and findings are more complex and nuanced than in fields like physics. By "boundaries of generalizability", I mean the extent to which a proposed model of the world makes accurate predictions across dimensions of entities, space, and time (Moeller et al., 2022). For example, Newton's laws of motion and gravitation built a model of the natural world that makes reliable and accurate predictions about everything from the trajectory of an apple falling from a tree to the paths of celestial bodies. The model's limits are only encountered when you start considering extremely small particles, extremely large masses, or extremely high speeds. The "volume of validity" contained in these boundaries of generalizability is massive.

By contrast, consider social cognitive theory (Bandura, 2000). What are its boundaries of generalizability? The results from Yeo and Neal (2006) might be considered evidence of at least one boundary condition. In their study on dynamic relationships between self-efficacy and performance on an air traffic control task, Yeo and Neal found a positive between-person relationship in self-efficacy and performance, as social cognitive theory would predict, but an opposite relationship existing at the within-person level. In the same way a physicist might zoom in on an object and find that the motion of its particles are better described by quantum theory than Newton's laws, Yeo and Neal found that resource allocation theory provided a better explanation for a specific activity and level of analysis than social cognitive theory.

Compared to Newton's laws, which have clearly defined limits that extend across the known universe, the range of potential exceptions to the general relationships described by social cognitive theory are many and complex. In physics, complexity in boundary conditions is considered a weakness of the theory, because its goal is to describe a universal underlying reality. For theories about human behavior, however, the less boundary conditions a theory has (i.e. the more universal it is), the more it necessarily simplifies the nature of the human experience. There is no one universal way to experience being human; individual differences are not deviations from a "universal underlying human" that we all approximate, they are simply differences (Bryan et al., 2021). Models of human behavior, therefore, must always navigate trade-offs between their generalizability and ability to make accurate predictions about specific humans, specific behaviors, and specific situations (Curren & Wirth, 2004).

A consequence of this is that when a result doesn't replicate, it may simply indicate a boundary of generalizability rather than a manifestation of the "replication crisis" (Renkewitz & Heene 2019). As Moeller et al. (2022) argues, a lack of replicability in tests of individual-, context-, and time-moderated theories like social cognitive theory shouldn't necessarily be interpreted as untrustworthy findings when there are a multitude of possible boundary conditions we potentially haven't recognized. In fact, as technology gives us the ability to measure smaller and smaller units of individual experience, context, and time, we might expect truly replicable results to become even more rare as they become more and more dependent on the unique "quantum effects" of specific individuals, in specific contexts and specific moments in time.

The challenge with this, of course, is that a multitude of possible boundary conditions makes it difficult to assess the value of a theory by way of isolated and disconnected hypothesis testing. When boundary conditions are underspecified, it becomes possible to justify the result of any empirical test as evidence of a potential boundary condition rather than a threat to the validity of the theory we are testing. For example, consider the results from the study I presented in

Chapter 2. I found evidence supporting social cognitive theory for independent reading activities, but not shared reading activities. The interpretation of this result depends on how the boundary conditions of social cognitive theory are defined. If social cognitive theory predicted a universal relationship across activities, these mixed findings would represent a potential threat to the validity of the theory. Instead, social cognitive theory defines self-efficacy contextually, so it is conceivable for the dynamics of parental self-efficacy to have different relationships to different types of parenting activities (Bandura, 2006). Therefore, our mixed results do not necessarily threaten the validity of social cognitive theory because they potentially reflect a boundary condition.

Any “failed” replication of Chapter 2’s study can be similarly justified. Say we repeated the study from Chapter 2 with a different group of parents and obtained a different result: that daily levels of parental self-efficacy were related to shared reading activities instead of independent reading activities. Does this refute our theory? Not necessarily, because it is entirely conceivable that this particular group of parents have different relationships in their parental self-efficacy and their child’s home literacy activities. What if we repeated the experiment with the same parents, but during a different 14-day window? Again, because the dynamics of self-efficacy may change based on context, any result can be explained by way of unmeasured contextual variability unique to this particular trial, rather than a threat to the validity of the theory.

At the end of the day, a small collection of disconnected hypothesis tests gives us very little empirical information about the range of conditions under which social cognitive theory will reliably make valid predictions. A single hypothesis test only illuminates a single point, a single confluence of possible conditions in the universe. A single point does not establish a boundary. Boundaries are defined by at least two points: one point on one side of the boundary, and one point on the other side of the boundary. But why stop there? With many tests that systematically explore different directions on across a potential boundary surface (e.g. via factorial experiments), we can illuminate the nature of more complex boundaries[1]. In order to do this systematically, however, we need lots of data, which is why researchers have begun advocating for the investment in shared research infrastructure to enable the collection of larger data sets (Bryan et al., 2021).

In Chapter 3, I demonstrated a way we can collect significantly more data in our evaluations of educational interventions using existing technology. By integrating measurement and implementation by way of digital exhaust and EMA approaches, I showed it is possible to dramatically increase the detail and intensity of our measurement designs in education intervention research without the traditional cost and burden. The pilot study I presented illustrated how mobile technology-based materials could be simultaneously used as a measurement platform to collect detailed and intensive measures of implementation and implementation-related processes in family literacy programs. Although I focused on the measurement of home literacy practices and intervention, the approaches I presented can be applied to collect more detailed and intensive data for use in the study of individual, contextual, and implementation processes in any educational context.

As we work to increase our data collection capabilities in educational intervention research, whether by way of investment in shared infrastructure, technology, or by other means, it is important we consider the ways different measurement designs and analyses make assumptions about the existence of individual differences (or lack thereof), and prioritize some forms of heterogeneity over others. Without a systematic way to organize the universe of possible individual differences that are possible to observe in educational intervention, we risk over- or

[1] Quantitative methods such as structural equation model trees (Brandmaier et al., 2016) and finite mixture models (McLachlan et al., 2019) can also aid in this discovery process, but a complete discussion of these methods are beyond the scope of this text.

underrepresenting types of heterogeneity that might otherwise play a critical role in our theories and their boundary conditions. In the next section, I present Cattell's (1952) data box, a theoretical framework that builds a somewhat exhaustive taxonomy of observable heterogeneity by which we can organize existing and future measurement designs in educational research.

Cattell's (1952) Data Box

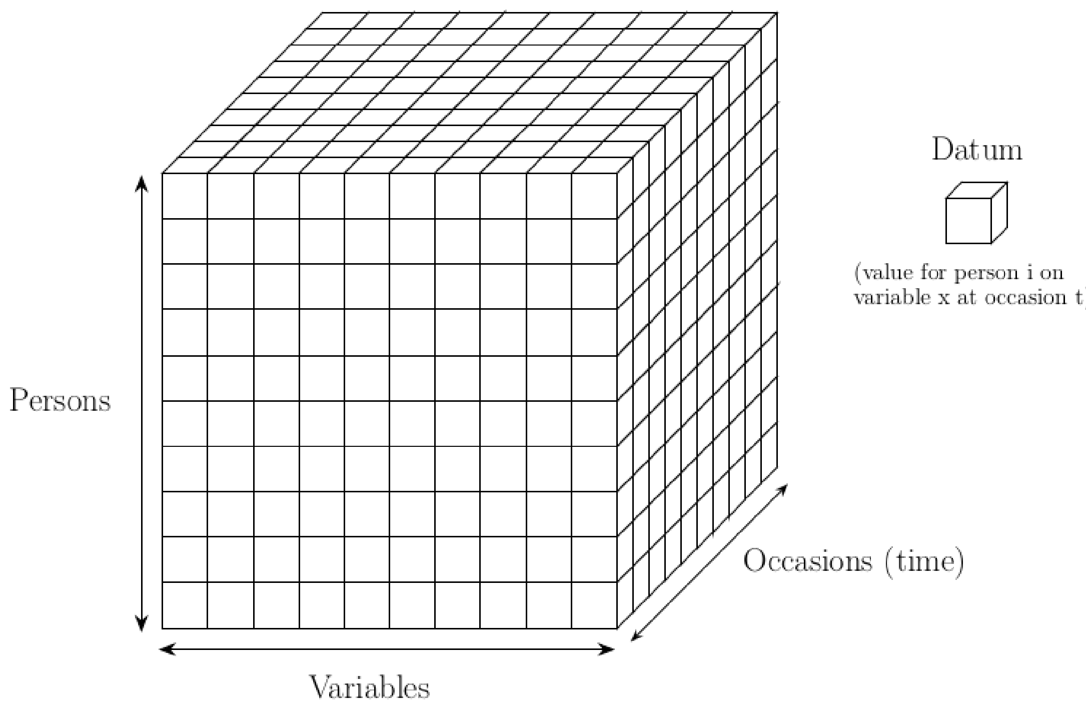

Figure 4-1: Cattell's (1952) Data Box (as illustrated by Ram & Nesselroade, 2007)

Cattell's (1952) "data box", and its later variants (e.g. Cattell, 1988; Ram & Nesselroade, 2007) provide a useful framework for organizing the types of variation that can exist in developmental systems. It is useful to apply in our study of measurement design in educational research, because all educational processes are, by definition, developmental processes.

Educational processes, like all developmental processes, are observed by measuring variables on samples of individuals at different occasions in time. The data box organizes these data in the form of a three-dimensional cube, with dimensions corresponding to persons, variables, and occasions. Each datum, as illustrated in Figure 4-1, represents a single measurement on a specific variable, for a specific person, at a specific moment in time. This representation of measurement is useful for illustrating how different forms of individual differences are captured (or not captured) by different measurement designs.

By taking cross-sections of the data box, we find representations of common measurement designs found in educational research. For example, a cross-section at a single occasion result in the familiar Person x Variables data matrix collected by a standard "cross-sectional", or "between-person" study designs in educational research (Figure 4-2). Similarly, a cross-section on a single person results in the Variables x Occasions matrix that characterizes data from a standard "single-subject" or "within-person" study design in educational research (Figure 4-3).

Experimental evaluations of educational interventions have often adopted pre / post measurement designs, represented by the data box as two Person x Variables matrices, one for each measurement occasion (Figure 4-4). The goal of educational intervention is to affect with-person change, which requires with-person measurement. We test for differences in a pre / post randomized controlled trial design by comparing the mean within-person difference in an outcome between groups of individuals receiving the control or treatment condition. This is not the only way to test for causal effects of an intervention, however: single-subject designs in educational research are increasingly making an appearance in educational research (e.g. U.S. Department of Education, 2017). An "ABAB" single subject experimental design (Kratochwill et al., 2021), for example, estimates the mean within-person difference in an outcome between groups of control and treatment occasions (Figure 4-5).

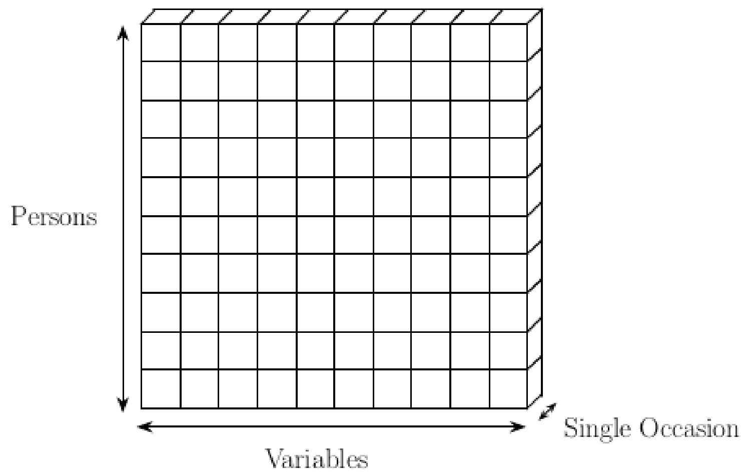

Figure 4-2: A standard “cross-sectional”, or “between-person” study design as represented by the data box

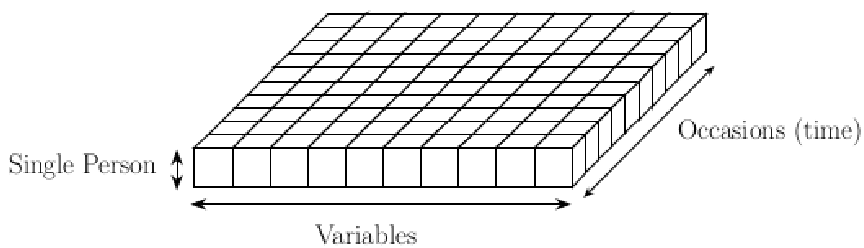

Figure 4-3: A standard “single subject” or “within-person” study design as represented by the data box

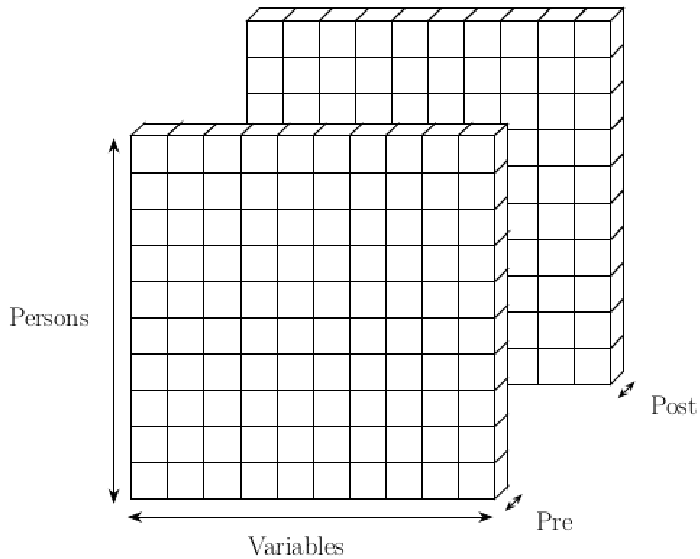

Figure 4-4: A “pre / post” study design

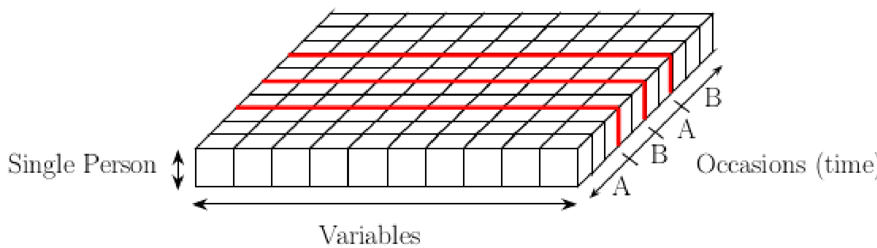

Figure 4-5: A single subject “ABAB” experimental design

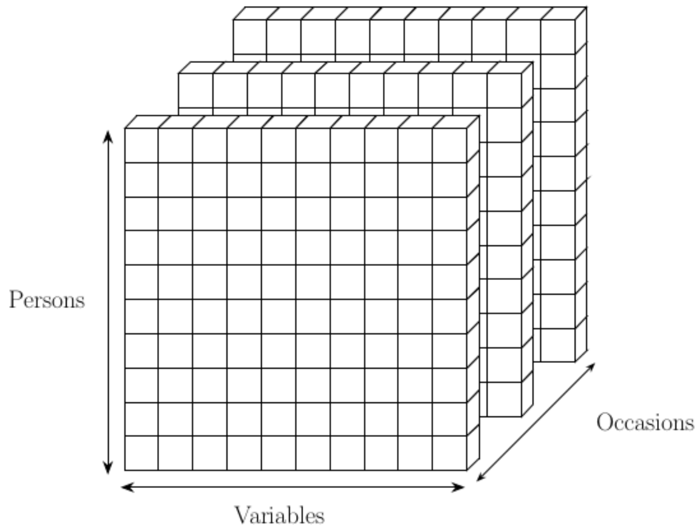

Figure 4-6: A longitudinal “panel” study design

Limits of Generalizability

Cattell’s data box is a useful tool to consider the boundaries of generalization in our findings, because it highlights which dimensions of heterogeneity are being considered in different study designs, and what is assumed to be homogeneous when making a generalization. In a standard cross-sectional observational study (Figure 4-2), when the group is representative of a population, we can expect relationships found in the group will apply to the larger population. Similarly, in a single-subject study design (Figure 4-3), when the sampled occasions are representative of typical moments in that person’s life, we can expect average relationships between variables in the collection of sampled occasions to apply to a larger duration of that individual’s life. For results found in a single-subject design to be applied to other individuals requires a strong assumption of homogeneity in a population. For results found in a cross-sectional study to apply to other occasions in time requires a strong assumption about homogeneity in time.

Observational studies that employ longitudinal measurement designs can help assess potential limits of generalizability in both axes simultaneously by measuring multiple individuals at multiple occasions, explicitly measuring differences between individuals in a sample, and differences within individuals across time (Figure 4-6). Panel designs, intensive longitudinal designs, and measurement “burst” designs all fall into this category (Trivellato, 1999; Bolger & Laurenceau, 2013; Sliwinski, 2008). This makes it possible to understand boundary conditions by way of statistical methods that simultaneously estimate individual and generalizable dynamics across time, such as Group Iterative Multiple Model Estimation (GIMME; Gates et al., 2020).

The data box indicates a similar counterpoint in between-group and single-subject experimental designs. In an “ABAB” single subject experimental design, more “AB” cycles randomized across time can increase our belief that the intervention had an average effect across the observed duration of time for that individual by balancing heterogeneous temporal effects. In a randomized controlled trial, more participants can increase our belief that the intervention had an average effect for that particular group at that moment in time by balancing heterogeneous individual effects. Likewise, a common criticism of single subject experimental designs is that they don’t “generalize”: that we can show a causal effect of an intervention in a single individual does not imply that it will work for other individuals. But an identical criticism can be made about the generalizability of randomized controlled trials, because observing an effect at one moment in time does not imply that it will work again in the future.

In educational intervention research, as in much of behavioral science and psychological research, we have historically focused on checking assumptions in order to make valid generalizations from small units to large units, but have been lax when it comes to assumptions about generalizations from large units to small (Molenaar 2004). For example, although randomized controlled trials tell you the intervention worked on average for the individuals in your sample, it cannot tell you that it worked for everyone in the sample without strong assumptions of homogeneity of the individuals in your sample. It is entirely possible that the intervention works for some, but not others. It might even be that the intervention works very well for some, but is iatrogenic for others. This is also true across the axis of time in single subject designs, although perhaps more self-evident: generalizing the results of an ABAB single-subject experimental design to specific occasions in time requires strong assumptions of homogeneity in an intervention's response across occasions.

Longitudinal experimental designs randomized across individuals and occasions can combine the strengths of group-level and single-subject designs. This can facilitate analyses of the limits of generalizability of an intervention's mechanism by enabling comparisons of the within-person effects of the intervention between individuals (or subgroups). Methods for analyzing these types of longitudinal data and experimental designs that perform randomizations in time as well as in populations are beyond the scope of the present discussion (for examples, see Klasnja et al., 2015; Collins et al., 2007).

Statistical methods aside, a key barrier to the wider use of longitudinal measurement designs in intervention evaluation is that they are resource-intensive. Traditionally, with a finite amount of resources, a project can collect a finite amount of data points; an increase in participants requires a decrease in occasions, and vice-versa. Strategic use of technology, as I described in the preceding chapters, can reduce the cost and burden of measurement, without these traditional limitations.

Extending Cattell's Data Box

So far, we have shown how to define an intervention's effect using the data box representing individuals. Now we consider how to extend the framework to represent different forms of heterogeneity in the context and implementation of interventions. Cattell himself offers an updated (1988) version of the data box in 10 dimensions; this representation is somewhat overkill for our purposes. Instead, we take the direction of Ram and Nesselroade (2007), by representing measures of "Context" as an extension of individual variables (Figure 4-7).

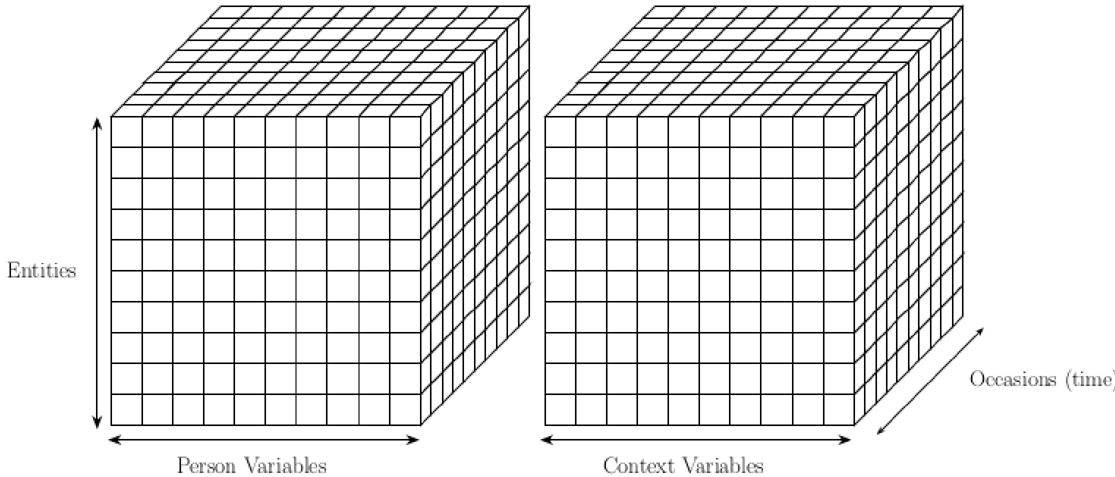

Figure 4-7: Extending the data box to include contextual variation

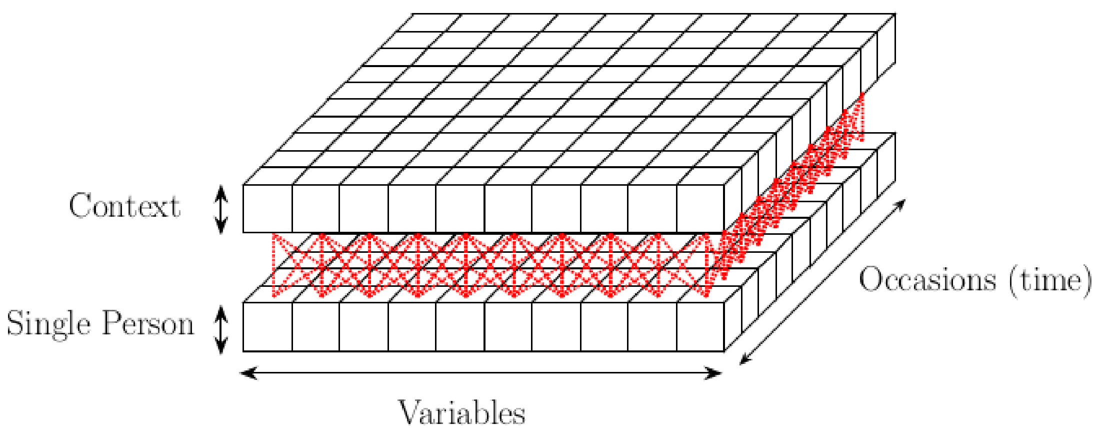

Figure 4-8: Visualizing Person x Context interactions (couplings between person and context in red)

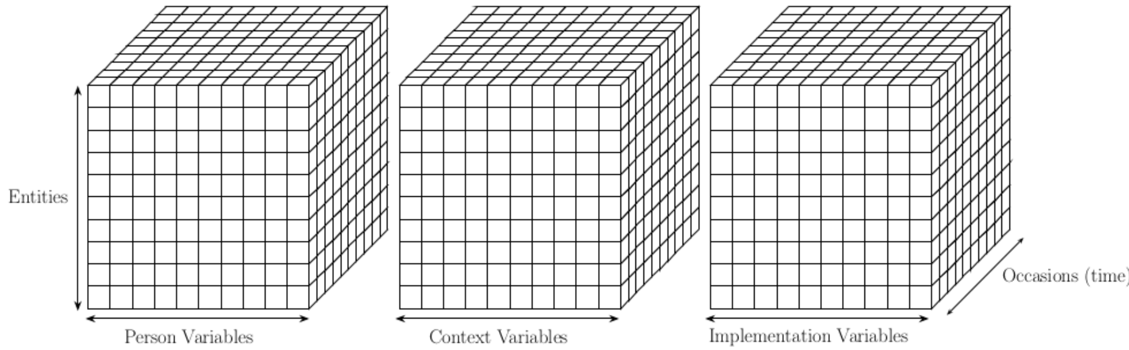

Figure 4-9: Extending the data box to include implementation differences.

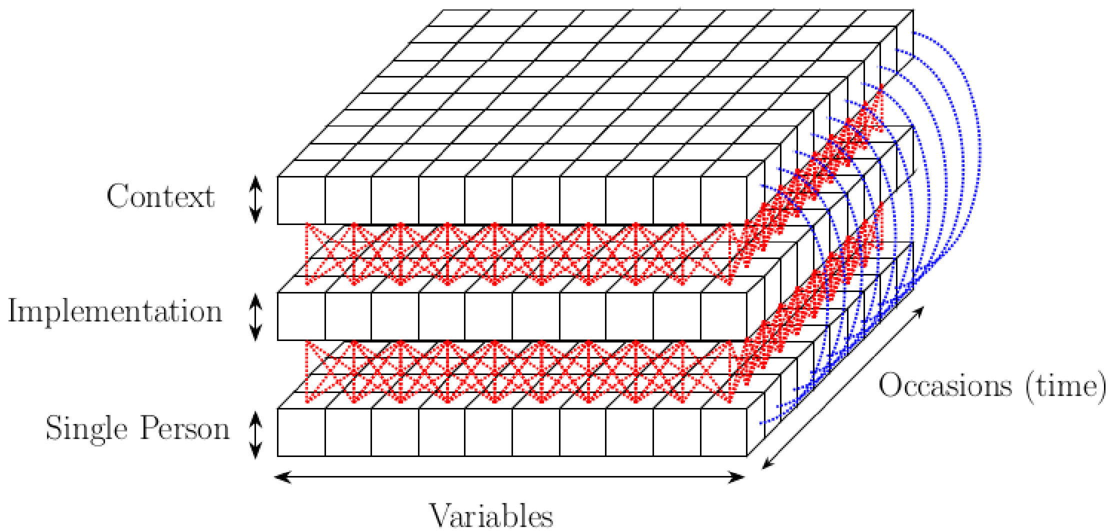

Figure 4-10: Visualizing Person x Context x Intervention interactions. Note that for representation purposes this diagram does not represent the exhaustive couplings between context and person.

As with individuals, we can consider differences between as well as within contexts across time. In education research, the differences between contexts we study often relate to physical locations or geographic areas, such as the differences existing in the demographics between two schools. This could also refer to more abstract notions of context, such as the difference between home and school education environments. As with individuals, contexts are not fixed; heterogeneity can be found within contexts as they exhibit change in their variables across time. For example, the demographics in a school may change over time, or a child's home and school environments may change in the middle of a school year if their family moves to a new location.

Process-level theories about educational processes consider bidirectional interactions between individuals and their environment, as represented in Figure 4-8. Obtaining measures of individuals and contexts as they change over time enables us to test theories about the bidirectional effects of their interactions. Just as individuals exhibit different behaviors across time, they also exhibit different behaviors across contexts. Ecological Momentary Assessment (EMA) study designs, such as the study I presented in Chapter 2, are a useful way by which we can collect measures about individuals' behavior as they move through contexts of interest.

Finally, we can add another data cube to our framework to describe the nature of heterogeneity found in intervention implementations, as shown in Figure 4-9. Just as we discussed with individuals and contexts, we can consider differences in implementation variables existing between interventions as well as within interventions across time. Differences existing between implementations might be represented in terms of specific intervention components or even specific procedural elements. For example, one implementation of a shared reading intervention might use a dialogic reading approach, and another might use a paused reading approach (e.g. Noble et al., 2020). Differences existing within an implementation refer to how it changes across time. These repeated measurements could represent fluctuations in its adherence or fidelity across time, or the pacing of elements delivered in its curriculum.

Just as we saw with individual and context, each observation of an intervention's implementation in a moment in time is coupled to the simultaneous state of the individual receiving the intervention and the context they are receiving it in. Figure 4-10 depicts a visualization of this three-way interaction. Just as individual and contextual processes interact and drive change in individuals and contexts across time, implementations of interventions can be considered processes in themselves, which interact with the interaction of individuals and contexts across time.

An intervention implemented in one way, in one context, with one group of individuals, at one time, does not give us much information about how variations in individuals, contexts, and implementations may change the outcomes of the intervention. Testing theory about “what works, for whom, under what circumstances” and characterizing its boundary conditions requires measuring and relating heterogeneity existing between and within “what” (implementations), “whom” (individuals), and “circumstances” (contexts). The data box helps organize these sources of heterogeneity in a manageable way by which we can better understand the assumptions being made in our measurement designs and analyses.

Concluding Remarks

In educational science, one of the primary ways we currently set boundary conditions in our tests of theoretical mechanisms is by focusing on population-level generalizability. Defining a target population is one way to set the boundaries of the universe under which you expect a theoretical mechanism to generalize. By carefully recruiting a statistically representative sample of individuals from the target population, we can use results from our sample to make statistical inferences about the larger population. With this approach, we can be reasonably confident that if we repeated our study on another representative sample of the population (or to the entire population at once), we will obtain similar average results (provided we also assume that the theoretical mechanism under study retains its validity across time and implementations of the study).

Although this approach can be very useful for estimating the effects of an education policy or intervention on a specific population of individuals, other designs are still required to additionally give us information about the range and nature of the conditions under which our theoretical mechanism makes valid predictions (Joyce & Cartwright, 2020). Although a representative sampling methodology enables us to generalize results to a larger group, that larger group still only represents a single unit, a single “point” of validity in an infinite universe of possible ways to form a group of individuals and contexts. In addition to individual and contextual considerations, the validity of this efficacy estimate for future implementations also depends on the extent to which future efforts exactly replicate its protocols, leaving little guidance for navigating local constraints. By contrast, if we design our evaluations such that they produce an understanding of the range and nature of the boundary conditions of the salient mechanisms in the intervention, we could provide education decision-makers with relevant information about “what works, for whom, under what circumstances”, to inform local adaptations (Gutierrez & Penuel, 2014; Weiss et al., 2014).

Furthermore, a focus on population-level generalizability without also investigating the boundary conditions that may exist for a theoretical mechanism within a target population can give us the false impression that a theoretical mechanism is equally valid for every subgroup or individual that makes up the larger group (Molenaar 2004; Bryan et al., 2021). This has dangerous implications for the potential applications of our findings. When the evaluation of an intervention’s effect estimates a positive effect for a population, this is only an estimate of its mean effect across all the individuals in the population. This does not necessarily imply that the intervention was positive for every individual in the sample: perhaps it had an iatrogenic effect for a minority group in the population but was well-received by a majority group. Or maybe we give up on an intervention that showed no population-level effects without realizing there was a small subgroup that found it extremely beneficial. Without careful consideration of the ways differences may exist within and between individuals and groups, we increase our risk of doing harm in the world and potentially miss opportunities where we could have had a positive impact.

Investing in our data collection capabilities is necessary to ensure our efforts to improve the world are resulting in positive and equitable effects (Joyce & Cartwright, 2018).

Recent advances in the fields of AI and data mining make a systematic study of boundary conditions all the more imperative. Methods from these fields can be used to build models that are highly predictive without providing understandable explanations of their predictions (Shmueli, 2010). Although efforts are underway to increase their interpretability (e.g. Gilpin et al., 2018), their limitations remain difficult to characterize and understand. It is not uncommon for these models to appear to make accurate and useful predictions, but still contain extremely harmful and pernicious forms of bias such as racism and sexism (Zou & Schiebinger, 2018). Extreme care will be required as we begin to consider incorporating these models into our educational decision-making in practice and policy (Leslie, 2019).

Because no model we develop will reflect the entire complexity of an individual's experience, all the models we develop for use in education science warrant careful humility, a continual checking and re-checking of our assumptions. If we wish to use our scientific understanding to drive equitable and inclusive decision-making in education policy and intervention, developing a nuanced understanding of the limitations of our theoretical models is just as important as developing the models themselves. But doing this requires systematic efforts to collect more data reflecting the many types of heterogeneity that can exist between and within individuals, contexts, and implementations of interventions and policy. As the studies presented in this dissertation illustrate, there's a lot of progress we can make in this direction with the technology and resources we already have.

References

Bandura, A. (1977). *Self-efficacy: Toward a Unifying Theory of Behavioral Change*.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research* (Illustrated edition). The Guilford Press.

Bryan, C., Tipton, E., & Yeager, D. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01143-3

Cattell, R. B. (1952). The three basic factor-analytic research designs—Their interrelations and derivatives. *Psychological Bulletin*, *49*(5), 499–520. https://doi.org/10.1037/h0054245

Cattell, R. B. (1988). *Its Ordering of Total Resources in Terms of Possible Relational Systems*. 62.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New Methods for More Potent eHealth Interventions. *American Journal of Preventive Medicine*, *32*(5, Supplement), S112–S118. https://doi.org/10.1016/j.amepre.2007.01.022

Curran, P. J., & Wirth, R. J. (2004). Interindividual Differences in Intraindividual Variation: Balancing Internal and External Validity. *Measurement : Interdisciplinary Research and Perspectives*, *2*(4), 219–247.

Gates, K. M., Fisher, Z., & Bollen, K. A. (2020). Latent Variable GIMME. *Psychological Methods*, *25*(2), 227–242. https://doi.org/10.1037/met0000229

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to Practice as a Criterion for Rigor. *Educational Researcher*, *43*(1), 19–23. https://doi.org/10.3102/0013189X13520289

Joyce, K. E., & Cartwright, N. (2018). Meeting our standards for educational justice: Doing our best with the evidence. *Theory and Research in Education*, *16*(1), 3–22. https://doi.org/10.1177/1477878518756565

Joyce, K. E., & Cartwright, N. (2020). Bridging the Gap Between Research and Practice: Predicting What Will Work Locally. *American Educational Research Journal*, *57*(3), 1045–1082. https://doi.org/10.3102/0002831219866687

Klasnja, P., Hekler, E. B., Shiffman, S., Boruvka, A., Almirall, D., Tewari, A., & Murphy, S. A. (2015). Micro-Randomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *34*(0), 1220–1228. https://doi.org/10.1037/hea0000305

Kratochwill, T. R., Horner, R. H., Levin, J. R., Machalicek, W., Ferron, J., & Johnson, A. (2021). Single-case design standards: An update and proposed upgrades. *Journal of School Psychology*, *89*, 91–105. https://doi.org/10.1016/j.jsp.2021.10.006

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. https://doi.org/10.5281/zenodo.3240529

Moeller, J., Dietrich, J., Neubauer, A. B., Brose, A., Kühnel, J., Dehne, M., Jähne, M. F., Schmiedek, F., Bellhäuser, H., Malmberg, L.-E., Stockinger, K., Riediger, M., & Pekrun, R. (2022). *Generalizability Crisis Meets Heterogeneity Revolution: Determining Under Which Boundary Conditions Findings Replicate and Generalize* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5wsna

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*, *28*, 100290. https://doi.org/10.1016/j.edurev.2019.100290

Ram, N., & Nesselroade, J. R. (2007). Modeling intraindividual and intracontextual change: Rendering developmental contextualism operational. *Modeling Contextual Effects in Longitudinal Studies*, 325–342.

Renkewitz, F., & Heene, M. (2019). The Replication Crisis and Open Science in Psychology. *Zeitschrift Für Psychologie*, *227*(4), 233–236. https://doi.org/10.1027/2151-2604/a000389

Sliwinski, M. J. (2008). Measurement-Burst Designs for Social Health Research. *Social and Personality Psychology Compass*, *2*(1), 245–261. https://doi.org/10.1111/j.1751-9004.2007.00043.x

Tipton, E. (2013). Stratified Sampling Using Cluster Analysis: A Sample Selection Strategy for Improved Generalizations From Experiments. *Evaluation Review*, *37*(2), 109–139. https://doi.org/10.1177/0193841X13516324

Trivellato, U. (1999). *Issues in the Design and Analysis of Panel Studies: A Cursory Review*.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2017). What works clearinghouse standards handbook (Version 4.0). *Washington, DC: Institute of Educational Sciences Retrieved from Https://Ies. Ed. Gov/Ncee/Wwc/Handbooks*.

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How Much Do the Effects of Education and Training Programs Vary Across Sites? Evidence

From Past Multisite Randomized Trials. *Journal of Research on Educational Effectiveness*, *10*(4), 843–876. https://doi.org/10.1080/19345747.2017.1300719

Yeo, G. B., & Neal, A. (2006). An examination of the dynamic relationship between self-efficacy and performance across levels of analysis and levels of specificity. *Journal of Applied Psychology*, *91*(5), 1088–1101. https://doi.org/10.1037/0021-9010.91.5.1088

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, *559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Appendix: Item-level Analyses of Decoding App Pilot Data

In Chapter 3, I presented a mobile app that automatically collected digital exhaust and EMA measures of FLP enactment. In my analyses, I focused on the ability of these measures to answer research questions about enactment-related processes. In this Appendix, I demonstrate how the item-level data collected by the app can also be used to study the development of decoding ability and test theories about processes related to children's performance during enactment activities.

Estimation of Latent Skill Trajectories

The Syllabee mobile app collected item-response data during enactment sessions by integrating into Syllabee's error-handling procedures. In the Syllabee FLP, parents are trained to not immediately correct their children when they misidentify a word. Instead, parents write their child's incorrect response next to the target word and engage in prompting to help their child uncover their own mistake. In the mobile app, this procedure was emulated by way of parents swiping in a particular direction when their children made a mistake, and then typing in their child's response for use in the error-handling procedure. In doing so, collecting these data became a part of the intervention, rather than as a separate step for capturing item-response data.

Figure A-1 shows the session-level item response accuracies of the three dyads during the pilot of the mobile app. As can be seen by the graphs, raw average response accuracy scores do not capture skill growth across time, because parents were instructed to manipulate the difficulty of the activities as their children increased in ability. Furthermore, even within a selected difficulty level, items from previous levels were intermixed to provide continual review. In order to use these data to estimate skill trajectories across time, item-level difficulties must be considered along with response accuracy.

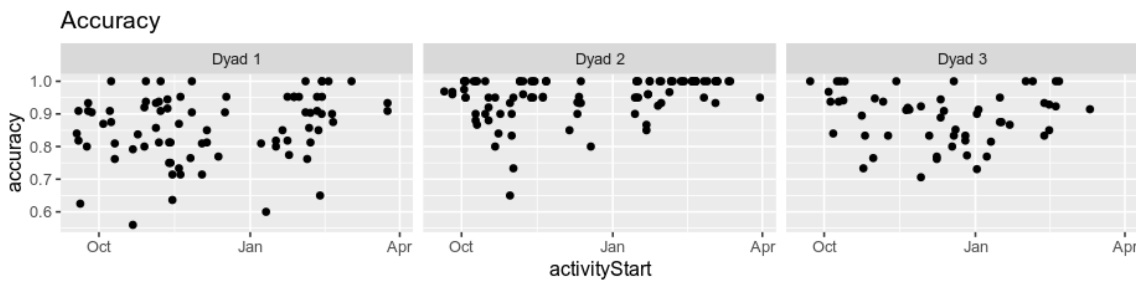

Figure A-1: Session-level aggregation of item-response data collected by the app.

The Elo Rating System

One method for estimating latent skill trajectories using the item response data collected by the Syllabee app's digital exhaust is the Elo rating system (Elo, 1978). The Elo rating system is a common approach for estimating latent skill from item-level response data in computer adaptive educational interventions (Pelánek, 2016; Klinkenberg et al., 2011). Originally developed to track player ability levels in the game of chess, the system has since been utilized by many games and sports. In the context of chess, players are given ability ratings that are updated based on the outcome of chess matches. Computer adaptive education interventions use the Elo system by reinterpreting individual "matches" as the outcome of item-level responses. When an individual answers an item correctly, the individual's latent skill estimate is increased, and the item's difficulty is decreased according to the difference between the current estimate of the individual's ability. When items are answered incorrectly, the reverse update occurs: the individual's ability estimate is decreased, and the item's difficulty estimate is increased.

More formally, given student s with current skill θ_s answering item i with difficulty d_i, the expected probability of a correct answer is given by the following logistic equation:

$$P(correct_{si}) = 1/(1 + e^{-(\theta_s - d_i)})$$

Based on the individual's actual response to the item, the individual's skill estimate and item difficulties are updated as follows:

$$\theta_s' = \theta_s + K(correct_{si} - P(correct_{si}))$$

$$d_i' = d_i + K(P(correct_{si}) - correct_{si})$$

Where $correct_{si}$ is 1 if the student answered correctly, and 0 if they answered incorrectly, and the initial ratings of skill θ_s and difficulty d_i are set to 0. The constant K is sometimes called the "K-factor", a hyperparameter that governs the size of adjustments at each step.

With our relatively small sample of individuals, it was not practical to attempt to estimate the difficulty level of every unique nonword. Instead, we fixed item-level difficulties according to their levels of introduction in the a-priori systematic phonics progression used by the Syllabee FLP. This difficulty progression categorized the difficulty of nonwords beginning with simple two-letter nonwords and gradually adding new symbol sound patterns and phonics rules. Table A-1 shows a selection of these levels. Figure A-2 shows a plot of latent skill trajectories obtained from the mobile app pilot data using the Elo rating method, with the K-factor tuned to produce monotonic skill trajectories.

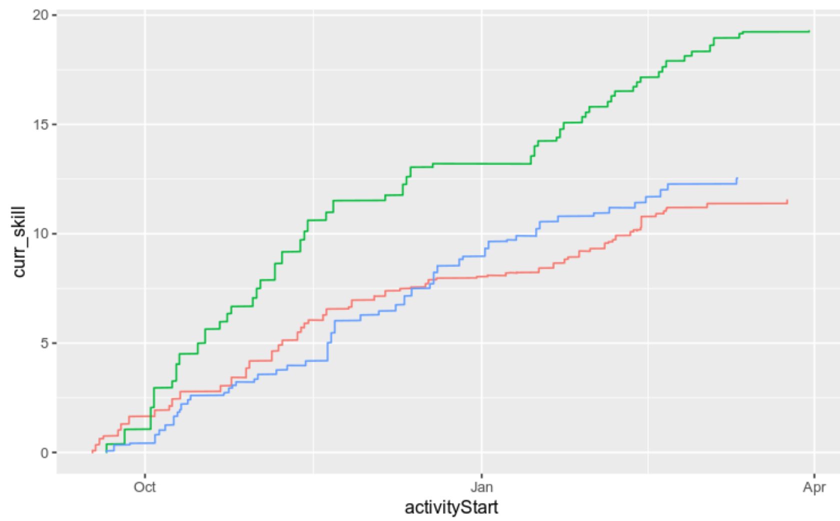

Figure A-2: Latent skill trajectories estimated using the Elo rating system.

Table A-1: Excerpt from Syllabee's word difficulty progression

Level	**Description**	**Examples**
1	VC (simple vowels, all closed)	it, om
2	CVC (simple vowels, all closed)	mit, zom
3	add b/d/p/g	pit, mod
4	"A's 3rd sound" (add ay, open vowels)	may, fa
5	"Intro Final E"	moke, five
6	"4 letter digraphs"	thad, mich
…	…	…
20	"One symbol, multiple sounds"	threan, trooms

Initial Skill Estimation and Trajectory Smoothing

Although the Elo rating system provides an estimate of current skill level that has shown to be useful for informing item-level adaptation (e.g., Klinkenberg et al., 2011), it does not update its previous estimates of skill as new information becomes available. This is readily apparent in

Figure A-2 by the fact that the initial skill level of each dyad is presumed to be equal. This limits the usability of past skill estimates in analyses that may wish to examine relationships in individual skill levels and other processes as they unfold across time.

One way to remedy this is by building a model of latent skill in state space, and then using existing statistical packages to estimate individual initial conditions along with smoothed skill estimates. As previous work has observed (Ingram, 2021), the Elo rating system algorithm approximates a steady-state Kalman filter. The Elo rating system algorithm implies the following system dynamics and measurement model in a state space representation of skill:

$$\theta_s[k] = \theta_s[k-1] + w[k]$$

$$logit\big(E(correct[k])\big) = \theta_s[k] - d[k] + \epsilon[k]$$

Where $\theta_s[k]$ and $\theta_s[k-1]$ are the current and previous estimates of the student's latent skill, $d[k]$ is the difficulty of the current item, $w[k]$ and $\epsilon[k]$ are normally distributed random variables representing latent skill innovations and measurement error respectively. For simplicity here we consider item difficulties fixed and known a-priori, but item difficulties could also be estimated simultaneously as additional state parameters. Formulated in this way, it is apparent that the K-factor hyperparameter in the Elo system is analogous to Kalman gain: a tradeoff between the latent skill uncertainty and confidence that the item response reflects the individual's true ability (Ingram, 2021). Figure A-3 shows the resulting skill trajectories estimated from the pilot data using the bssm package in R (Helske & Vihola 2022), with Kalman smoothing and initial skill level estimation.

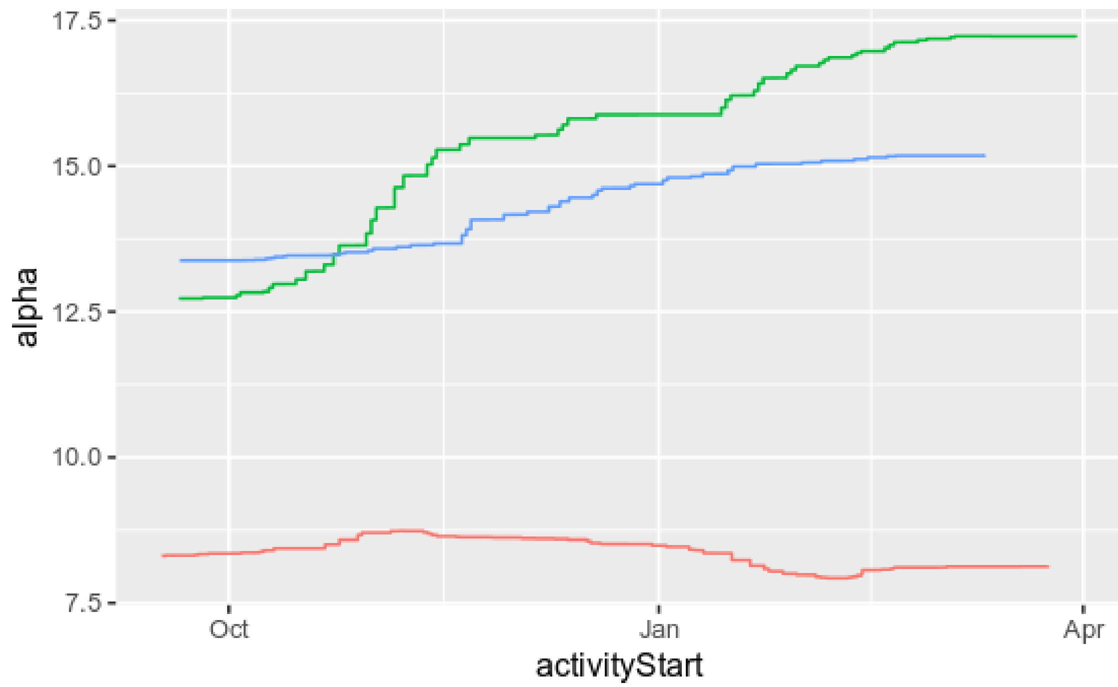

Figure A-3: Smoothed trajectories with initial conditions estimated using the bssm R package (Helske & Vihola 2022).

Relationships Between Affect and Performance

An individual's performance at a given moment is not solely determined by cognitive ability, but is also influenced by social, emotional, and other contextual processes (Bronfenbrenner & Morris, 2006). Previous theoretical and empirical work suggests that positive affect and arousal will have a bi-directional effect on cognitive performance (Ashby et al, 1999; Ashby et al., 2002). The individual estimates of latent skill trajectories obtained in the previous section give us the ability to control for an individual's skill level in a present moment, and test for relationships between children's emotional experiences during literacy activities and their correct responses. We hypothesized that children's session-level morale and energy levels would be positively related to their item-level response correctness.

To test these hypotheses, we used a generalized linear model to predict the probability of an individual answering an item correctly from the individual's present skill level, the item's difficulty, and their session-level morale and fatigue rating. The results are shown in Table A-2.

Table A-2: Results from a generalized linear model predicting the logit-transformed correct response probability from child morale and energy.

	logit(P(correct))
(Intercept)	-7.58 (2.19)*
Child Morale (session-level)	0.33 (0.06)*
Child Energy (session-level)	-0.17 (0.06)*
Covariates	
Current skill	0.36 (0.08)*
Word difficulty	-0.14 (0.02)*
Child Morale (mean)	2.42 (0.69)*
Child Energy (mean)	-0.05 (0.31)
AIC	3050.72
BIC	3096.20
Log Likelihood	-1518.36
Deviance	3036.72
Num. obs.	4905

*$p < 0.05$

In line with our hypothesis, we found that child morale (emotional valence) was positively related with the probability of answering a given item correctly, even when controlling for current skill level and item-level difficulty. Interestingly, we also found that child energy (emotional arousal) was negatively related with the probability of answering items correctly, contrary to our initial hypothesis.

Our finding that higher levels of child energy were associated with decreased levels of performance may be because the higher levels of arousal were related to higher levels of distraction for the child, rather than an increased focus on the decoding activities. Another possibility is that increased levels of arousal were related to children's levels of anxiety, which has been shown to reduce cognitive performance as well (Eysenck & Calvo, 1992). To test this possibility, we added a term to the model to see if children's emotional valence moderated the effect of their emotional arousal on their performance, but we unfortunately lacked the power in our small sample to establish this relationship. A larger study would be necessary to test for these interactions, as well as determine the generalizability of the present results.

Conclusion

Technology-enhanced intervention materials have the potential to collect enormous amounts of data via digital exhaust and EMA without traditional tradeoffs in cost and burden. In this appendix I demonstrated how item-level response data collected during a decoding focused FLP could be used to form estimates of an individual's skill trajectory across time. I showed how these estimates could be used along with other EMA measures to test theories about processes related to the development of an individual's decoding ability across time. Although the methods presented here show promise for studying the development of decoding ability and its interaction with other social, emotional and contextual processes, future work should consider assessing their external validity by comparing estimates of skill trajectories to existing standardized measures of performance using a larger sample of individuals.

References

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*, 529–550. https://doi.org/10.1037/0033-295X.106.3.529

Ashby, F., Valentin, V., & Turken, A. (2002). *11. The effects of positive affect and arousal on working memory and executive attention* (Vol. 44, pp. 245–287). https://doi.org/10.1075/aicr.44.11ash

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of Child Psychology*.

Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children*, *52*(3), 219–232. https://doi.org/10.1177/001440298505200303

Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.

Eysenck, M. (2012). *Attention and Arousal: Cognition and Performance*. Springer Science & Business Media.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Helske, J., & Vihola, M. (2022). *bssm: Bayesian Inference of Non-Linear and Non-Gaussian State Space Models* (2.0.1). https://cran.r-project.org/web/packages/bssm/index.html

Ingram, M. (2021). How to extend Elo: A Bayesian perspective. *Journal of Quantitative Analysis in Sports*, *17*(3), 203–219. https://doi.org/10.1515/jqas-2020-0066

Park, J. Y., Cornillie, F., van der Maas, H. L. J., & Van Den Noortgate, W. (2019). A Multidimensional IRT Approach for Dynamically Monitoring Ability Growth in Computerized Practice Environments. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00620

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers & Education*, *98*, 169–179. https://doi.org/10.1016/j.compedu.2016.03.017

Proust-Lima, C., Philipps, V., Perrot, B., Blanchin, M., & Sébille, V. (2022). Modeling repeated self-reported outcome data: A continuous-time longitudinal Item Response Theory model. *Methods*, *204*, 386–395. https://doi.org/10.1016/j.ymeth.2022.01.005

VITA

Kyle D. Husmann

Education

The Pennsylvania State University, University Park, PA

Ph.D. Human Development and Family Studies — August 2023

M.S. Human Development and Family Studies — 2019

California Polytechnic State University, San Luis Obispo, CA

B.S. Electrical Engineering, Valedictorian — 2011

Fellowships and Grants

Training Interdisciplinary Educational Scientists (TIES) Fellowship, US IES	2018
University Graduate Fellowship, Penn State University	2017

Awards and Honors

Douglas Research Endowment Award	2021
Hintz Fellowship Award	2021
Honorable Mention, NSF Graduate Research Fellowship Program (GRFP)	2019
1st Place, College-wide Rapid Research Competition	2018
Ford Foundation Scholarship	2017

Selected Peer-Reviewed Publications

Husmann, K. D., Brick T. R., DiPerna J. C., (2022) Applying the Measurement Model of Derivatives to Evaluate and Refine Measurement Scales in Longitudinal Education Data Using the ECLS-K. Journal of School Psychology.