1 Measurement Burst Designs and Peer Relations

One of the core factors that researchers and practitioners in developmental psychology are concerned with is the concept of change and variability. Longitudinal research methodologies allow scientists and practitioners to investigate how psychological phenomena change or are maintained across a determined period of time. However, in studying change across the lifespan, we must be cognizant of the variability that exists in how and when we measure psychological constructs in general. For example, in the study of internalizing symptoms among children, elevated levels of test anxiety, which refers to the negative internalized feelings experienced when confronted with academically evaluative situations (Spielberger & Vagg, 1995; Sub & Prabha, 2003), might be present due to an upcoming or previously taken exam. Social anxiety differs to test anxiety such that the former is typically associated with negative thoughts prior to, during, or after a social evaluative situation (Hearn et al., 2017), while the latter is more salient to test-taking experiences. Given these differences, levels of test anxiety are not likely to reflect the genuineness of an adolescent's anxious feelings toward tests in general, but rather a time-dependent or "snapshot" moment. One way to address changes and fluctuations in psychological states is via the use of measurement burst designs. With this in mind, we propose that the measurement burst would allow for increasing the accuracy with which we study phenomena common to youth development. Specifically, we focus on peer relationships, a fundamental context that becomes increasingly influential as children become adolescents. In three separate studies, we examine the extent to which the measurement burst design provides more-stable estimates of (1) adolescent experiences with being accepted/liked within the peer group, (2) social and test anxiety, and (3) the general self-concept, which refers to the evaluation of one's own competence to engage, maintain, and experience positive social interactions with others (Harter, 2012).

1.1 Longitudinal Research as Integral to Studying Development

Developmental and quantitative psychologists, including Paul Baltes (1987) and John Nesselroade (1990), have long argued that change and variability at the level of the individual is complex, multidirectional, and multidimensional, hence the need for rigorous measurements. Longitudinal research designs, in which measurements are taken at predetermined points across time, chart the development of the construct or phenomena of interest. These assessments are typically taken only once and then compared with the time points that follow (see Figure 1a). The results, depending on the length of the study, are multiple repeated measures of data points that chart the trajectory or path of a given

construct. For example, a study by Eisenberg and colleagues (1999) investigated the development and stability of prosocial dispositions by observing a sample of children from around the age of four or five years until early adulthood. The result was approximately eleven observations across this period of time, which demonstrated that individuals who possessed early prosocial personality dispositions (e.g., engaged in spontaneous sharing in childhood) had higher levels of prosocial behaviours in early adulthood. This widely cited study demonstrates the value of longitudinal research in charting change and variability across multiple developmental periods.

Clearly, longitudinal studies are valuable in demonstrating the degree to which change and variability occur. However, there are four main limitations to note. First, conducting longitudinal research can be financially costly and labour-intensive. Expenses directed to recruitment, maintenance, and completion of a study can contribute to increased costs with designing a longitudinal study. Second, as with other research designs, the risk of participant attrition and burnout is high. A study as intensive as that of Eisenberg et al. (1999) originally consisted of thirty-seven children but obtained a final sample size of thirty-two. The loss of five participants may not seem like a significant one, but it can be particularly harmful for longitudinal research in which participants are asked to be involved for months or even years. Similarly, a meta-analysis assessing the effects of cognitive behaviour therapy on chronic pain found that attrition rates in samples of children can range from 0 to 54 percent (Karlson & Rapoff, 2009). The cost-benefit of being involved for so long, coupled with trying to take advantage of the longitudinal design in order to measure a number of different constructs, can increase participant burnout, consequently increasing the likelihood of attrition. As a result, high attrition rates can have detrimental consequences for the generalizability of a study's findings (e.g., Gustavson et al., 2012). Regardless, the information obtained from longitudinal research plays a fundamental role in developmental research; consequently, methods to prevent participant burnout and attrition should be valued in our research.

The third limitation reflects on the measurement process more broadly. Whereas the Eisenberg et al. (1999) study was able to use multiple annual time points to measure prosocial dispositions, other studies have shorter time frames to work with, resulting in fewer observations. For example, a longitudinal study of video gaming assessed children and adolescents across three time points and observed that greater amounts of video gaming, among other factors, were risk factors for becoming pathological gamers (Gentile et al., 2011). The difference between these studies reflects an important limitation, in that having so few measurements, namely one at each time, can be influenced by other circumstances that occur at that given time. For example, the time at which

video-gaming frequency was measured could be influenced by the fact that children had less work to do at that specific time. Relatedly, when developmental researchers measure self-concept among youth, an adolescent's self-reported high score on a single-time assessment of their positive self-concept might be inflated because they received compliments on their appearance or performed extremely well on a test. Thus, that positive self-concept at the level of the adolescent may not be indicative of their true sense of self or what they feel. As such, it can be argued that while longitudinal designs can be useful for assessing changes in trait values, they do not allow researchers the opportunity to capture "systematic time- or situation-specific 'ups and downs' in individual's true state scores around the fixed trait" (Geiser et al., 2015, p. 2).

The fourth limitation reflects participant burden. Individuals who conduct research with humans and other animals face multiple challenges. One challenge is project completion. After assembling the needed funds, one needs to put the plans for the study into action. The basic criteria for success include having a sufficiently large sample that is representative of the target population and having a minimal amount or proportion of missing data. A further criterion for a longitudinal study is maintaining sample membership across the times of assessment. A second challenge concerns data quality. Our measures need to reach well-known psychometric standards of validity and reliability. To reach these stringent expectations, we need to use measures and procedures that will produce well-focused measures and that minimize measurement error. A third challenge concerns the ethical commitments researchers make to the participants in our studies and to the contexts where their projects occur. Researchers make a promise to the participants in their studies that their participation will not be a source of stress greater than what is "normal" in their daily lives. They also make a commitment, either explicitly or implicitly, to minimize the intrusive disruptions they bring to the contexts where their studies take place. These ethical commitments are a basic duty of research scientists.

These four challenges are situated in different domains. The first has to do with procedural issues related to obtaining and maintaining a proper sample. The second concerns the adequacy of measurements. The third is about ethics. Despite the differences between these challenges, each of them is linked to a critical but often overlooked concept known as participant or respondent burden. Participant burden refers to the demands placed on individuals who take part in a research study (Sharp & Frankel, 1983). This concept is not new. Discussions of it have been seen in the health and social science literatures for 100 years (Chapin, 1920). We conceptualize participant burden as having a curvilinear rather than a linear association with time.

Due to this time-related exponential increase in participant burden, the increased burden of a twenty-minute assessment period compared with a fifteen-minute period will be smaller than an increase from twenty minutes to twenty-five minutes when studying other constructs.

Participant burden has ethical and pragmatic consequences (Lingler et al., 2014). The ethical consequences are related to our commitment that participation in our studies will not be a source of stress. When participant burden is high, we violate this commitment. This violation occurs at the level of the individual participant and at the level of the contexts where we conduct our studies. At both levels, we do not live up to our agreement to treat research participants in an ethical manner. The pragmatic consequences of participant burden have to do with the quality of the data that are collected and with the continuity in participation in longitudinal studies. To the degree that participant burden causes fatigue and distraction, it will undermine a participant's motivation and ability to provide honest and accurate answers. These conditions will decrease the reliability and validity of our data. Reduced validity and reliability have an adverse effect on the adequacy of our studies. Another negative consequence is increased levels of attrition. Participants who feel burdened by their initial experiences in a study are unlikely to continue. Given that this form of attrition is not likely to be random, it will decrease the representativeness of a sample. It will also decrease the size of the sample. Both of these effects reduce the value of our studies. That being said, we recognize that true measurement burst designs can increase participant burden, depending on the intensity of the assessment, such as having participants provide emotional state ratings multiple times per day for thirty days. However, it can also help to ease the burden when it is incorporated as part of larger studies. Here, we posit that a measurement burst design can contribute by minimizing participant burden. Our point is that having two or more abbreviated assessment sessions instead of one very long session will reduce the demands we place on our participants and, as a consequence, will improve the ethical standards of our projects and will increase the quality of our data. In the following sections, we demonstrate how participant burden is reduced by having two testing sessions within each burst, which lasts between thirty and forty-five minutes. This would be in contrast to having a single session that lasts between sixty and ninety minutes, which is problematic. For example, having such long sessions significantly reduces class time for teachers and can increase the cognitive load imposed on students to focus on questionnaires that researchers ask them to take seriously. As such, multiple sessions that are shorter in length help to ease these burdens.

1.2 Measurement Burst Design Methodology

The measurement burst design methodology is an intensive form of assessment in which two or more measures are made in quick succession within time points, otherwise known as a burst (Sliwinski, 2008). Specifically, it provides a means of detecting *intraindividual* variability and change in one's mood and related emotions and behaviours by using multiple short-term measurements in a longitudinal research design (Nesselroade, 1990; Sliwinski, 2008). For example, in a three-wave longitudinal study that measures the stability of anxiety across the school year, each burst would consist of two (at minimum) assessments or observations, separated by one week, resulting in six observations for each participant (see Figure 1b). When multiple observations are taken within the same time point, we can increase the precision with which we are measuring the phenomena of interest by considering momentary deviations in state mood levels that can occur from one time point to the next as a result of external factors beyond the control of the researcher. In doing so, the high degree of measurement overlap within burst designs should produce a more stable estimate across time because of the increased accuracy of the two measures of the same construct. Specifically, a measurement burst design helps researchers understand nuanced changes in an individual's self-reported state at different points in time.

Sliwinski (2008, 2011) describes measurement burst designs as a way of capturing two extreme forms of longitudinal research designs. Indeed, the measurement burst design is a mix between longitudinal designs in which there are more widely spaced intervals and those in which data are collected in quick succession (e.g., daily diaries, experience sampling methods/ecological momentary assessments). The former captures more broad changes obtained at one point in time, while the latter allows for more fine-grained analyses of phenomena that may be prone to increased variability, such as affective states and self-esteem (e.g., Nelis & Bukowski, 2019). Importantly, the measurement burst design can help to disentangle whether the differences observed across bursts are typical ups and downs in one's mood versus an enduring change in one's trait mood over time (Geiser et al., 2015). As such, the measurement burst design offers a wide array of advantages over and above the two extremes of longitudinal data collection, including improved measurement of various phenomena (Stawski et al., 2016). For example, measurement burst designs allow for multiple, intensive measurements that increase the precision of the construct (e.g., anxiety) but also the reliability of the instrument being used. In addition, they also afford researchers the opportunity to maximize the data collection process by collecting as much information as possible while reducing

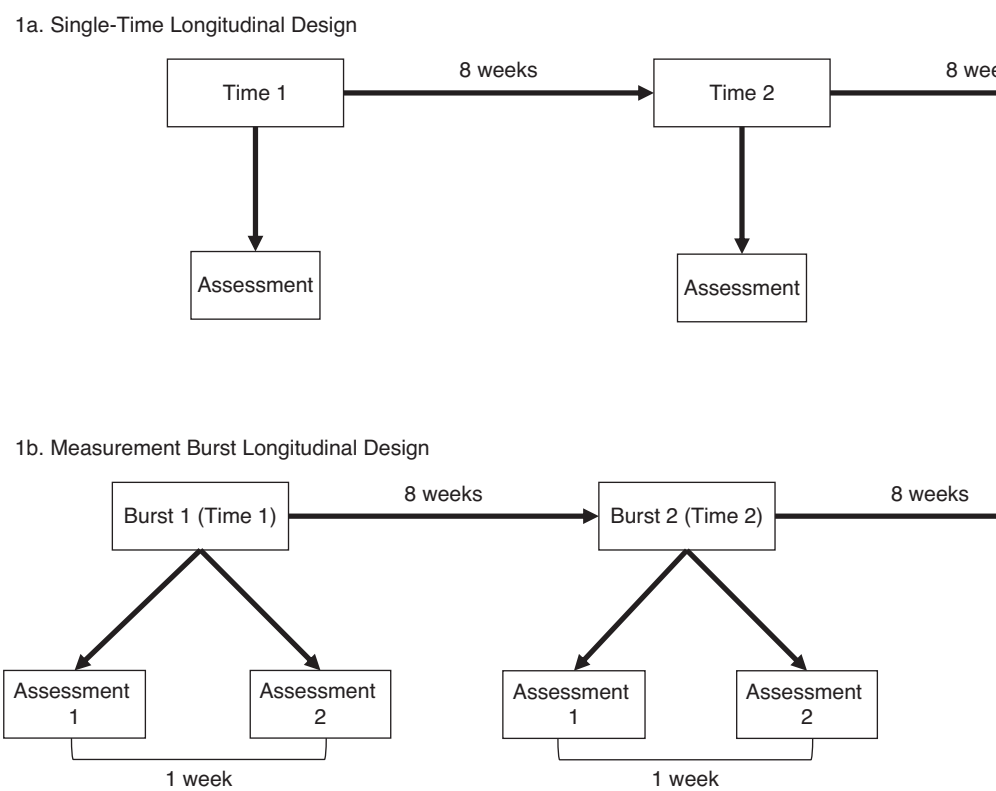

Figure 1 Comparison between the standard longitudinal design and the measu

participant burden and fatigue. Moreover, measurement bursts allow researchers to study the dynamics of change over multiple time points, namely within the bursts and over a longer time period, which is the overarching goal of developmental science (Sliwinski, 2008; Stawski et al., 2016).

The general approach of a measurement burst design study involves three main steps, two of which involve deciding on the temporal sampling of data within and across time points. Rast and colleagues (2012) argue that different patterns of variability emerge as a function of the construct under study. Typically, the longitudinal aspect ranges in months or even years. In our studies, we used three time points (known as bursts) within an academic school year. Time 1 (T1) was shortly after the beginning of the school year. Time 2 (T2) occurred approximately eight weeks later, which was followed by Time 3 (T3) that took place around eight weeks after T2. These time points were chosen to reflect the uncertainty that typically comes with entering a new grade at the beginning of the school year and at two points in which they should, in theory, be more comfortable with and knowledgeable about their class. The second step was to decide how often to collect data (i.e., assessments within bursts). Although there does not appear to be a set of rules that dictate how many and how often assessments within bursts should be conducted, there should be multiple assessments done in quick succession. This would suggest that the interval between assessments be short, such as days or a week. Another consideration to be made is the time investment on the part of the participants as well as other stakeholders. The goal of researchers should be to make the best use of the time participants take to give high-quality responses, and to do so by minimizing burden. In our studies, we had to consider the time allotted to us in the classroom. In this case, to minimize participant burden and maximize the valuable time afforded to us by teachers, each burst had two assessments, one week apart. The third step involves the methodological approach. In their chapter, Cho and colleagues (2019) highlight studies that demonstrate the utility of measurement bursts using various approaches, including daily diary data (e.g., Almeida et al., 2009) and ecological momentary assessments (EMAs) whose range can vary between minutes, hours, and days (e.g., Liao et al., 2017), and combinations of both (e.g., Scott et al., 2015). Each assessment type evaluates an individual's self-reported state and experience, which can then be used to chart change and variability.

A review of studies that apply measurement burst designs shows that they are often used with adult populations, studying change and variability in the physical, cognitive, and social-emotional domains (e.g., Lee et al., 2018; Scott et al., 2015; Sliwinski et al., 2010). However, there is evidence for its use with younger samples as well (see Riediger & Rauers, 2018 for a review of experience sampling in developmental research). For example, a daily diary

study examined the effects of students' trait self-regulation and perceived task difficulty on task enjoyment and independence from their parents during home-schooling as a result of the COVID-19 pandemic (Blume et al., 2020). Parents reported on their children's learning independence, perception of task difficulty, and enjoyment once a day for twenty-two days. There were three main findings to note on the within-person and between-person levels. First, there were positive associations between trait self-regulation and learning independence. Second, learning independence was found to be higher on days when tasks were seen as being easier, whereas on days in which tasks were difficult, learning independence was lower. Third, higher average task enjoyment was associated with greater learning independence, suggesting that those who enjoyed their task were also more independent than those who did not enjoy it. Studies such as the one by Blume and colleagues support the utility of measurement burst designs within the context of child development and functioning.

To our knowledge, the use of measurement burst designs is still emerging in peer research. Whereas cross-sectional intensive measurement designs are not new, incorporating them into longitudinal measurement burst designs is indeed novel. For example, a study by Lehman and Repetti (2007) explored the effects of negative school (e.g., academic problems) events on children's functioning using daily reports across five days. Their results showed that on days when youth reported greater problems in the academic or peer domain during the school day, they were more likely to report aversive interactions with their parents, suggesting that peer experiences contribute to variability in children's functioning. More recent work by Schmidt and colleagues (2020) examined associations between relatedness satisfaction and relatedness frustration at school on child-reported positive and negative affect. Participants reported on these measures twice daily for two weeks (Study 1) and then once a day for four weeks (Studies 2 and 3). Broadly, their findings showed positive associations between relatedness satisfaction and positive affect, and between relatedness frustration and negative affect at the between-person level. Moreover, relatedness frustration positively predicted negative affect, suggesting that variations in children's reports of having less-than-positive encounters with peers were associated with greater negative affect. They also showed within-person level findings on the positive effect of relatedness satisfaction on positive affect, indicating that children who had positive interactions with their peers demonstrated more positive affect. Though cross-sectionally designed, these studies represent initial attempts to incorporate measurement bursts into peer research, which indeed support its utility, despite requiring further investigations.

Given the degree of change and stability in various facets of a child's or an adolescent's life, the measurement burst design nested within longitudinal

research can help to increase the reliability of the measures being evaluated. With this in mind, this Element aims to provide evidence for the application of burst designs within the domain of peer relations research among early adolescents in three different ways. In particular, our goal is to demonstrate that measurement bursts offer more-accurate estimates than single-time estimates on constructs commonly studied in peer research. In Section 2, we demonstrate the effects of the burst in sociometric ratings and nominations of acceptance, which are fundamental measures within peer research. In Sections 3 and 4, we focus on self-reports of core psychological concepts related to early adolescent experiences with internalizing problems (i.e., social and test anxiety) and the self-concept (or self-worth). In these sections, we hypothesize that the stability between the respective measures will be stronger when the burst model is applied, in comparison to typical longitudinal designs in which one assessment is made at each time point.

1.3 Peer Relations as a Developmental Context

Social relationships are part of the fabric of the human experience. The relationships individuals have with others can significantly impact the well-being and overall functioning of everyone involved. Peer relationships are commonly seen as a particularly salient context through which youth development can be studied. Research on peer relations in childhood and adolescence extends to the early days of experimental psychology (Bukowski et al., 2018; Monroe, 1898). The interactions that youth have with their peers represent some of the crucial social relationships that play a role in one's well-being and development. Unlike family relationships, peer relations are unique in that they are voluntary and are more balanced in terms of power (e.g., Howe et al., 2011; Laursen & Bukowski, 1997). As such, peers offer a unique insight into how extrafamilial relationships can affect the developmental trajectories of children and adolescents.

Many studies within the peer domain tend to use the longitudinal design framework to examine various correlates and consequences associated with peer relations. For example, the *Child Development Project* is a longitudinal data set that began when children were five years old and followed them through adolescence, which includes approximately nine waves of measurement (see Dodge et al., 1990; Pettit et al., 2001). Findings using this data set have shown several different results related to peers. Specifically, Criss et al. (2002) found that grade 1 children's level of peer acceptance moderated the positive association between experiences with externalizing behaviours in the second grade, suggesting that positive peer experiences could be protective against individual stressors. Another data set, known as the *Waterloo Longitudinal Project* (see

Hymel et al., 1990), investigated youth development within the peer context using school-age children beginning at around five years old. In this case, Rubin and colleagues (1990) found social withdrawal to be stable during early school years and that it was positively associated with depression, feelings of loneliness, and negative self-perceptions of competence in late childhood. These findings tend to remain consistent across many studies on peer experiences; namely, those who experience greater peer rejection at one time show greater levels of externalizing (Hymel et al., 1990) and internalizing problems (Prinstein et al., 2005), at least three years after the initial data collection. Taken together, the enduring nature of peer research has allowed for the study of youth development from childhood to adolescence and has provided unique insights into the role of peers. However, as important as these longitudinal designs are to understanding change and variability in the peer context, the single-measurement approach has limited our ability to fully establish accuracy in what we are measuring within and across time.

Peer research often includes a variety of ways in which indices of acceptance, rejection, and individual well-being are measured. Two of the most common methods used among adolescents include sociometric assessments and self-reports. Sociometric assessments are the standard method of studying one's experience with peers (Bukowski et al., 2017; Cillessen & Bukowski, 2018). Typically, acceptance and liking within the peer group is measured via nominations of a participant's limited or unlimited choice of rank-ordered "same-sex" and "other-sex" friends (Bukowski & Newcomb, 1984; Coie et al., 1982; Newcomb & Bukowski, 1983). A follow-up assessment then asks participants to rate how much they like their classmates on a five-point Likert scale ranging from 1 ("not at all") to 5 ("like very much"). From there, nomination scores and liking ratings are produced, accounting for various factors including class size. Currently, there are multiple recent and comprehensive manuscripts that describe these procedures in detail, including a special issue by Marks (2017), as well as manuscripts by Bukowski et al., (2017), Cillessen and Bukowski (2018), and Velásquez et al. (2013). Clearly, sociometric assessments are the most broadly used method for evaluating how liked and accepted children and adolescents are within their peer group.

The second common method is through self-reported assessments of well-being and functioning. This is especially the case for adolescents, who are more capable than younger children of precisely describing their feelings, experiences, and thoughts on a variety of constructs, including anxiety (Wood et al., 2017), friendship (Persram et al., 2021), and self-concept (Nelis & Bukowski, 2019). By adolescence, cognitive gains generally include the ability to engage in self-evaluations and social comparisons with one's peers. Given