

Appendix A

Literature Review and Proposed Approach

Part I - Introduction - Why EMA, Why Now 1

Part II - Background On Proposed Virtues

DILIGENCE

- Conceptual Definitions 4

- Enhancers and Inhibitors 6

- Operationalization and Measurement 7

HONESTY

- Conceptual Definitions 9

- Enhancers and Inhibitors 10

- Operationalization and Measurement 11

COMPASSION

- Conceptual Definitions 12

- Enhancers and Inhibitors 16

- Operationalization and Measurement 18

Part III - Our Proposed Approach

DILIGENCE

- For Students 20

- For Teachers 22

- Measures for Validation 24

HONESTY

- For Students 25

- For Teachers 27

- Measures for Validation 27

COMPASSION

- For Students 28

- For Teachers 30

- Measures for Validation 31

Part I - Introduction - Why EMA, Why Now:

Look at most any school's mission or core values statement on its website and there is inevitably mention of at least a handful of character domains such as respect, empathy and honesty¹. Most schools proclaim as part of their mission the development of these traits in their students. Yet almost no school has the in-house knowledge and expertise to accurately measure whether it is successfully doing this. As a result, schools often fall back on dubious evidence for proving that they develop character in their student body, including writing anecdotal stories about student "moral exemplars" in their parent newsletters, requiring retrospective assessments of various student behaviors on 3 point scales (i.e. "Excellent," "Satisfactory," "Needs Improvement") every two to three months on progress reports, or simply pointing to the mere existence of character education programs in their curricula as de facto "proof" that they are developing character. None of these approaches come close to giving an empirically valid picture of what is really going on.

We contend that the necessary first step before designing and then implementing programs to promote character in schools is to figure out how to reliably measure character itself. The state of the art in the field of character assessment is disappointingly limited, even in university-led studies, let alone schools; most studies over-rely on trait-based questionnaire measures that require participants to answer long surveys reflecting on global statements (i.e., "I am a hard worker" (Duckworth & Peterson, 2007) or "I get upset when I see a girl being hurt" (Bryant, 1982)). There are also pervasive issues with any self-report measures when used with children, regardless of what type of questions are asked. For one, participants' lack of verbal ability and comprehension, especially when having to reflect on emotional states, can sometimes skew results (Zhou, Valiente & Eisenberg, 2003).

We propose to measure students' character by leveraging Ecological Momentary Assessment methodology, a technique of data collection that has shown great promise in recent years in social science research (Weinstein et al, 2007; Weinstein & Mermelstein, 2007; Tong et al, 2005;

¹ See Appendix C-1 for a table compiling the results of our content analysis of 50 mission statements from independent and public school from around New England.

Axelson et al, 2003). Our vision is to push a small number of daily questions to students and teachers about that day's behavior, moods and attitudes using smartphone-based apps. Recent data on the veritable explosion in smartphone usage worldwide (Orcutt, 2012) leads us to strongly contend that this is the direction most schools must go in to stay relevant in a quickly changing world.

Whereas early attempts at EMA required participants to fill out paper questionnaires at the sound of a beeper (Larson et al, 1990) or verbally answer questions on a phone call with a researcher (Axelson et al, 2003), smartphone technology can allow for much less time-consuming and invasive procedures. Once participants have downloaded the app and logged in with their username and password, they just respond to the "new question" alert by tapping an answer and submitting.

We also are intrigued by the possibility of using EMA measures of character because they, if done properly and with scientific rigor, can finally start to pry away at the mysteries of context in this domain of human behavior. Hartshorne and May's groundbreaking but ultimately discouraging study of character in children (1928) concluded that there was so little consistency across context in the traits they studied (such as honesty) that it was futile to conclude much of worth. This lack of consistency in moral character has finally been integrated into a promising conceptual model that takes into consideration "mixed character traits" (Miller, 2013) but little empirical work has been completed to date attempting to apply this framework.

Perhaps discouragingly, there have been consistent calls for many years in character research for measures that consider context beyond that of a psychology lab. A representative quotation, from Peterson & Seligman (2004), states:

"Do laboratory perseverance effects mirror real-world perseverance? ... Perseverance tests in the lab generally measure behavior over a duration of perhaps half an hour, and the setbacks involved are relatively benign. Can researchers safely assume that the predictors of lab persistence also predict real-life perseverance?" p. 246

This problem of context is paramount, and one that we believe we can alleviate with our proposed data collection techniques. Referring directly to this issue in her recent EdWeek article on character development in schools, Joan Goodman states:

"...it may not be good policy to try to make students gritty, for grittiness is probably situational, not a pervasive characteristic. While it seems intuitive to view people as possessing character traits (Sam is hard-working—when problems arise, he perseveres; Samantha is lazy—she loses concentration when faced with difficulties), in fact, a long history of research has failed to validate the possession of such traits. That is, a child's responses are not predictable across varying circumstances. Whether Sam will actually cheat on a test depends on whether he perceives it to be fair, desires to do well, is adequately prepared, and has available opportunities, as well as whether others cheat, the state of class and group morale, and so on." (online article, no page number)

Yet to rely solely on behavioral measures of character is equally problematic. Research on the Big Five trait of Conscientiousness, for example, has shown that there is no way of reliably accounting for possible past neurological damage in lab study participants, making it hard to know what a score on this trait actually reflects (Roberts et al, 2012). It has also been proposed that behavioral measures requiring some degree of previous learning can put people with lower IQ or education levels at disadvantages that could skew results (Buelow & Suhr, 2009).

As a result of both self-report and behavioral measures having significant drawbacks, researchers continue to call for some way of bridging the gap between the contextualized information gleaned from behavioral measures of character and the broad, trait-based information gathered from self-report surveys (Roberts et al, 2012). Our approach will attempt to do this by utilizing EMA reports over a long time period, an entire school year, rather than just a few weeks, which most studies using this methodology have done. This, we hope, will provide enough data to allay concerns about too little context to infer traits (Roberts et al, 2012).

Part II - Background On Proposed Virtues

Diligence/Grit

Conceptual Definitions:

In common language usage, the word diligence (from the Latin *diligentia* for "care, attentiveness" Origin - Old French: *carosse de diligence* – a fast stagecoach used for long journeys) has a range of meanings, from diligence as the legal minimum effort required for a task (i.e. due diligence) to an enduring characteristic of a person who consistently manifests industrious, caring and assiduous effort in both thought and action. When operationalized, however, most researchers measure diligence over short timeframes, like completing a mentally taxing task in a psychology lab experiment (see Peterson & Seligman, 2004 for review) and examine only short-term costs that may interfere with diligence such as tedium or embarrassment (Williams & DeSteno, 2008). A more comprehensive term, "grit," widens the scope and refers to diligence applied to long-term, relatively unchanging goals of personal relevance (Duckworth et al, 2007). For the purposes of this research we will retain the word "diligence" to denote the construct with the understanding that its definition includes the expanded sense of diligence captured by the word "grit". While recent excitement in education circles has been rightly tempered by cautious academics who worry where these virtues can lead someone without underlying moral aims (Goodman, 2012) it's understandable why we would want to raise children who can persist in the short term and achieve meaningful goals in the long term despite difficulties.

It's important to note, however, that diligence is conceptually different from simple "task persistence," a domain studied extensively in the literature (Peterson & Seligman, 2004). While "task persistence" refers to doing something difficult for an extended period of time (certainly an important part of any diligent effort) the term "diligence" connotes being careful and attentive as well. This points to related aspects of personality like the Big Five domain of Conscientiousness, which includes sub-facets like orderliness, formality and responsibility (Roberts et al, 2012) and research has shown consistent correlations between conscientiousness and diligence for long term goals (Duckworth & Peterson, 2007).

We argue also that this disposition can refer to awareness of the details of one's surroundings and

immediate experiences that doesn't necessarily require action. For instance, a mother carefully watching her three-month old child to make sure she doesn't harm herself could be considered "diligent," yet to an observer, she may look like she's doing very little.

Another important clarification must be made about whether the subjective opinion one has of a task changes whether the action being performed is considered diligent. Can it be considered "diligence" if someone persists with a boring and repetitive task because he is receiving only extrinsic rewards (like monetary compensation or course credit) but finds the task disagreeable and with no immediate relevance to his personal goals in the short-term? We would argue that it can.

Most cancer researchers could tell us that certain aspects of their work in the lab are not intrinsically rewarding and even disagreeable but the greater goal of curing cancer necessitates diligent work on these tasks. And referring back to our hypothetical mother and child, it's conceivable that the diligence she showed toward her daughter was not rewarding in and of itself in the short term, but nonetheless brought intrinsic value to her because she could successfully play the role of a parent able to protect her child from harm. Also, if an assembly line worker shows careful attention to his work each day only because he's getting paid a salary to do it, and as a result, the products being manufactured work properly for the consumer, this should also be considered an example of diligence. "Diligence" connotes careful, persistent effort or attention to tasks, regardless of whether they are intrinsically or extrinsically worthwhile. It only becomes something other than diligence once the effort lacks care and attention.

"Distress tolerance" is an important related domain to diligence, defined as "one's ability to persist in goal-directed behavior in the context of emotional distress" (Roberts et al, 2012, p. 8). But this domain should be considered only a sub-facet of diligence, one that heightens it but does not subsume it. An airplane mechanic who can tune out the constant barrage of engine noise in his day-to-day work, for example, may still do his work in a careless or haphazard manner, leading us to conclude that he exhibited a lack of diligence. Conversely, it is easy to think of an example of someone who was careful and persistent in his work on a task even though it caused him no emotional distress and actually gave him great joy. A professional baseball player honing

his swing in batting practice during the off-season with no immediate pressure to improve his results could be an example of this.

Enhancers and Inhibitors:

There is extensive research on the psychological and situational factors that lead to higher levels of diligence. Personality characteristics that augment diligence include curiosity (Kashdan & Fincham, 2002), self-control and self-discipline (Corker et al, 2012; Duckworth & Peterson, 2007; Peterson & Seligman, 2004), ability to focus attention and avoid distraction (Ommundsen et al, 2005) and ability to delay gratification (Mischel et al, 1989). Beliefs and perspectives that have been shown to heighten diligence and grit include possessing an internal locus of control (Hollenbeck, Williams & Klein, 1989), having an attributional style that explains the causes of one's negative experiences in terms of external factors rather than personal failings (Seligman & Schulman, 1986), believing one has competence in relevant skills (also known as "self-efficacy") (Trautwein et al, 2009), believing that goals should be pursued to reach mastery rather than gain acclaim from others (Peterson & Seligman, 2004), and believing intelligence is flexible rather than fixed from birth (Ommundsen et al, 2005). Situational factors that lead to higher rates of persistence include knowing that one's effort will be rewarded, being able to choose the task, and experiencing early successes along the path to completion (Peterson & Seligman, 2004; Kashdan & Fincham, 2002). Lastly, being in a good mood while pursuing a task tends to stave off the urge to quit (Peterson & Seligman, 2004).

In summary, diligence refers to dispositions that prioritize persistent attention and care to tasks that often, but not always, bring rewards that can either be intrinsic or extrinsic. These dispositions also require the ability to overcome internal resistance and external obstacles and setbacks on the path to completion.

Operationalization and Measurement:

Researchers primarily operationalize diligence in three broad ways:

1. Self- and other-reports of effort and behavior
2. Time spent on a difficult task
3. Care and attention given to a difficult task

Global self-reported effort put into schoolwork (i.e. "I really work hard on classwork assignments in mathematics") has been used (Trautwein et al, 2009) as well as reports of effort on specific assignments (i.e. "I put a lot of effort into preparing for this exam") (Corker et al, 2012). Observer reports of effort, such as from friends, coworkers and teachers, have also been used successfully in conjunction with other measures (Roberts et al, 2012).

Numerous global self-report scales have also been developed to assess factors that affect levels of diligence, such as learning and study strategies (Corker et al, 2012; Ommundsen et al, 2005), beliefs about one's ability to persist (Duckworth & Peterson, 2007; Lufi & Cohen, 1987), goal orientation (Peterson & Seligman, 2004) and tendency toward procrastination (McCown & Johnson, 1989). There is also at least one self-report scale that asks questions about how often someone performs specific behaviors associated with conscientiousness, of which some relate to diligence (Jackson et al, 2010).

Time spent on difficult tasks has been used as well (DeWall et al 2011; Williams & DeSteno, 2008) although it's important to note that not all of these studies asked participants about the level of interest or personal value these tasks held for them. As a result, persistence on these tasks may have also been conflated with simple allegiance to authority or anxiety about avoiding failure. Common ways to measure this have been counting the number of complicated math problems or word scrambles that were completed in a timed session (DeWall et al, 2011) or measuring how long participants spend on a "mental rotation" task requiring high levels of concentration before choosing to quit (Williams and DeSteno, 2008).

Level of care taken in one's work has been measured by asking participants to do 3-digit by 3-digit multiplication problems without use of a calculator or playing the classic children's board game *Operation*, in which a buzzer sounds any time a participant's tweezers touch the sides of the hole from which objects must be removed (DeWall et al, 2011). Some research has taken this

a step further and asked participants to complete a task correctly while being subjected to emotional distress (Roberts et al, 2012).

There seems to be no standard set of measures used for measuring diligence, although the sheer number of global self-report scales available in the relevant research literature (Peterson & Seligman, 2004) indirectly points to this approach as the most common, if not the most effective. The Grit Scale (Duckworth & Peterson, 2007) has been used extensively in the KIPP network of public charter schools in the past decade, and has so far been consistently correlated with academic and life success (Tough, 2012). The Persistence Scale for Children (Lufi & Cohen, 1987), while not cited extensively in subsequent research literature, is at least one of the few self-report measures of persistence that is not subsumed under a longer multi-faceted personality inventory (Peterson & Seligman, 2004). The Behavioral Indicators of Conscientiousness (BIC) (Jackson et al, 2010) utilizes the Act Frequency Approach to measurement and asks participants to reflect on how often they exhibit specific behaviors, from "never performed the behavior" to "perform the behavior quite often." Most of the items do not relate to diligence, but the Laziness and Industriousness subscales align very closely with our understanding of the domain, including "Persist at tasks after meeting setbacks or failures" and "Watch TV or go on the internet instead of taking care of responsibilities."

Lastly, there have been some methods employed that attempt to reduce recall bias by asking people to reflect on their behaviors from a single day. These methods include diary entries on self-perceived effort put into each night's homework (Trautwein et al, 2009), online surveys conducted each night asking whether specific behaviors that relate to conscientiousness were performed that day (Jackson et al, 2010), and asking participants to fill out surveys five times per day on a Palm Pilot device asking how well different adjectives described them ("How well does 'persistent' describe you in the past hour?") (Fleeson, 2001). These all seem to hold promise, if not in the exact content of all items, but in their approach that stresses reporting on one day's experiences. We would argue, however, that these measures are limited in their effectiveness as a result of being cumbersome and time-consuming. We propose a research design that we believe will reduce the momentary effort to respond significantly.

Honesty

Conceptual Foundations:

The virtue of honesty has sometimes been subsumed under that of integrity (Blasi, 2005), but we view it as simply a closely related domain. We limit our discussion of honesty with regard to two behaviors that we believe are most relevant to school contexts: telling the truth and avoidance of cheating. Honesty, then, is the disposition to tell the truth, when appropriate, in a variety of situations, in an appropriate way, and for virtuous motives (C. Miller, personal communication, August 2013). Cheating refers to any behavior that "intentionally breaks the relevant rules in a situation (whether moral or non-moral) in order to gain an advantage using deceit or fraud." (Miller, 2014, p. 54).

A comprehensive review of the research literature on honesty has been conducted elsewhere (Miller, 2014; Miller, 2013a) and for that reason we will not duplicate that effort here. The important thing to consider for us is the empirical conclusion that there is little consistency across contexts with regard to honest behavior, and therefore, we cannot state that all people exhibit the personality trait of honesty.

That said, a working definition of honesty must be discussed here, and to that end we lean heavily on the aforementioned work of Miller (2014; 2013a) who we believe has provided as comprehensive a depiction of various features of the trait of honesty as we can find. A paraphrased and blended account of his work on lying and cheating appears directly below. We believe this is best suited for use in a school environment where complex moral philosophical inquiry is not the norm:

An honest person will:

1. Refrain from telling lies, both "big" and "small," for self-oriented reasons (e.g., fear of embarrassment or anticipated material loss) except when morally justifiable (e.g., protecting the life of innocent Person A by lying to Person B who is trying to kill him);

2. Refrain from telling lies for purely hurtful reasons unless morally justifiable (e.g., telling Person B that he is a worthless weakling in order to buy time for Person A to escape harm);
3. Refrain from cheating (or allowing others to cheat) in situations where there is the freedom to do so and the rules are fair, even if he (or the other persons) could gain personal benefit by cheating.
4. When lying or cheating (or helping others to do so) in ways that are clearly morally wrong, internally confront his behavior by refraining from self-deception or rationalization, and feel subsequent remorse and attempt to refrain from similar behavior in the future.

In simpler language, a "perfectly" honest person tries to avoid exhibiting dishonest behavior, but if he has to, never does it for hurtful reasons, only does it for morally justifiable reasons, refrains from doing it (or enabling it in others) solely for his own benefit, recognizes he did a "bad thing" when he *does* act dishonestly, and feels subsequent guilt and/or shame for these transgressions, even if he is never caught.

Enhancers and Inhibitors:

Motivations to lie and cheat show a great degree of overlap with one another with only a few minor differences, so we will discuss them in terms of one overall grouping. Research on these behaviors (Mead et al, 2009; Mazar, Amir, Ariely, 2009; DePaulo & Kashy, 1998; Newstead et al, 1996; see Miller, 2013a Chapter 3 and 2014 Chapter 10 for a review) has shown that the main factors involved in whether or not someone exhibits these behaviors are:

- Avoiding embarrassment and/or shame;
- Avoiding loss (material, financial, of a relationship, etc.);
- Avoiding punishment and/or blame;
- Avoiding hurting someone's feelings;
- Wanting to hurt someone else;
- Because it was the morally correct thing to do;
- To maintain one's belief that he/she is honest;
- To gain competitive advantage or achieve goals of worth to the self or others;

- Because it seems thrilling or fun to do so;
- Being too tired to maintain self-control;

Note that, save for the last factor, each of these things could be either a motivation to be honest (i.e., it would hurt someone's feelings less if you told them the truth than lied) or to be dishonest. A better way to describe these factors, then, would be to say they are both enhancers and inhibitors of dishonest and honest behavior, depending on the situation (Miller, 2013a).

Operationalization and Measurement:

Because one of the empirically verified factors as to why people choose, or do not choose, to lie and cheat in various situations is to maintain a positive self-image with regard to honesty (an unequivocally moral domain), we think it is important to look at measures of "moral identity" (also referred to as "moral self") as part of our discussion of honesty. Moral identity has been defined as the degree to which a person perceives herself to be a moral person committed to ethical principles (Blasi, 2004). We argue that someone who considers moral principles, including honesty, to be more important to his sense of self will tend to exhibit less dishonest behavior across contexts. Some measures ask self-report questions regarding how much one feels that moral traits are central to one's identity (Aquino & Reed, 2002) and others ask about a mixture of principles and personality facets, some moral (honesty, generosity) and some not (athletic, logical) (Jimenez et al, 2008; Barriga et al, 2001; Arnold, 1993). There are also more complex, narrative assessments in which participants are prompted to describe themselves in various ways ("actual" or real self, "dreaded" self, "ideal" self) and then these paragraphs are coded for moral descriptors by experimenters (Power, 2006).

As for the virtue of honesty itself, the vast personality assessment literature has one widely validated inventory, the HEXACO-PI (Lee & Ashton, 2004) that directly assesses a facet they label Honesty-Humility through four subscales for Fairness, Sincerity, Greed Avoidance and Modesty. Honesty, not surprisingly, has also been assessed by asking questions about frequency of past cheating behavior (McCabe, 2008) and providing opportunities to cheat in experimental

situations and then gathering information on who chooses to do so (see Miller, 2013a for review; Mead et al, 2009; Mazar, Amir & Ariely, 2008; Casey & Burton, 1982).

Lastly, we are intrigued by an approach taken by DePaulo & Kashy (1998; See Miller, 2013a for detailed treatment) who utilized daily "lie diaries" that asked adolescents to record every social interaction and lie they told each day. They were also instructed to record whether they felt distress during and after the telling of lies, and what reasons they had for each lie they told (i.e., to avoid embarrassment, to avoid hurting the other person's feelings). This closely aligns with our philosophy of character assessment utilizing daily snapshots of behavior and attitudes to uncover crucial contextual factors.

Compassion

Conceptual Foundations:

The concept of love in human cultures has fascinated people for as long as history has been recorded (Peterson & Seligman, 2004), and the plethora of approaches in the disciplines of philosophy and psychology for categorizing various forms of it is an apt reflection of its arguably confounding nature (Hendrick & Hendrick, 1986; Lee, 1973; Lewis, 1971; Fromm, 1956). The Greeks had four words for love that ranged from passion between romantic partners to unconditional, altruistic caring for others, and this has often been the place where theoretical work begins. Yet it's surprising to see how much research on love defaults to discussion of either romantic relationships (Hendrick & Hendrick, 1986; Reis & Aron, 2008) or attachment relationships that have the early bonds formed between parent and child as their foundation (see Peterson & Seligman, 2004 for a review).

Our role as education consultants working with schools to better assess their students' character development necessitates that we confine our work to forms that are more feasible within these settings. These are the more altruistic forms of love that are exhibited between friends, community members and humanity at large that the Greeks referred to as *agape*. One central element of this form of love involves an emotion of concern for what is good for others, for their

own sake (C. Miller, personal communication, August 2013). Terms that are more palatable to a school and parent audience that have been used interchangeably (sometimes wrongly!) to describe this form of love are "empathy," "sympathy," "compassion," and "altruism." These are the words we will turn to now, and we will argue that only the concept of "compassion" fully encompasses enough aspects of caring and attention to warrant measuring in schools.

The domain that has been researched most extensively from this group is empathy, and in our work with schools this is the term most often used by school leaders to describe the aspects of character related to "positive feelings directed toward another person" (Storm & Storm, 2005, p. 336) that lead to desirable outcomes in school communities. It is a catch-all term that schools tend to find agreeable, yet based on the common definition used by most researchers, it seems empathy would not encapsulate all that a school is looking to instill in its charges. Most conceptual definitions in the research literature describe it fairly narrowly as a state of emotional arousal in response to the comprehension of another's emotional state (Zhou, Valiente & Eisenberg, 2003) and refer to vicariously feeling the emotions of another (Eisenberg & Miller, 1987). In simpler language, it is "feeling an emotion with someone, and because the other is feeling it" (Snow, 2000, p. 66). A further distinction often made is that empathy requires a form of perspective-taking where a person either imagines how he would feel in another's situation (known as "projective empathy") or imagines how the other would feel ("empathy proper") (Miller, 2013a).

It should be noted that none of these definitions stipulate what kind of emotions a person must feel along with another for it to be considered empathy. One can share uplifting as well as sad or painful emotions with another and still be considered to be feeling empathy, even though we would argue that in school settings, situations where a community member is feeling sad or in pain are the most crucial times in which others would hopefully feel empathic toward him or her.

Furthermore, it is easy to think of a situation in which feeling empathy for another might not lead to positive or "prosocial" action toward that person. It is one thing to feel sad alongside Sally because her father suddenly passed away; it is another thing to act in order to help relieve or soothe her suffering. It is true that extensive research literature on empathy has shown that it

very often *leads* to prosocial behavior (Eisenberg, Fabes & Spinrad, 2006; Batson, 1991, 1990; Eisenberg & Miller, 1987) but, technically speaking, it is only the first step toward the larger end of harmonious human relations. Empathy is, in short, the feeling that often, but not always, precipitates generous, kind, and caring actions. It can be considered a potential "enhancer" of helping behavior (Miller, 2013a). We believe, therefore, that it is not a useful term to encapsulate the virtue of love.

"Sympathy" is a closely related term described by most researchers as almost identical to "empathy," and frequently grows out of feelings of empathy (Eisenberg, Wentzel & Harris, 1998), but is an other-oriented emotion involving care and concerned attention for another person in need (Miller, 2013; Prinz, 2012; Batson, 1991). It likely comes out of a combination of shared empathic sadness and referring to previous knowledge of either the person for whom one feels sympathy or other people who were in similar situations (Eisenberg, Wentzel & Harris, 1998). This is an emotion directed toward another, not felt along with another, and does not require accurately perceiving what the other is experiencing (Miller, 2013a). Furthermore, whereas empathy is an "imaginative reconstruction" of someone else's experience, independent of judgment about it being good or bad (Crisp, 2008, p. 234), sympathy connotes an inherent judgment that what the other is experiencing is causing that person some form of suffering. Therefore, if sympathy involves a judgment that might not even come out of vicariously feeling an identical (or at least similar) emotion to another in need, this could lead to behaviors that come out of misdirected pity or condescension and be quite unhelpful to the other person in certain situations. Similar to the limitations surrounding the concept of empathy, we argue that sympathy is also not adequate to describe all that is entailed by the word "love." Even if we limit our discussion to non-romantic love between friends, family, community members and strangers in the wider world, the virtue involves not only feelings of concern and care, but an impetus to give selflessly in an altruistic manner.

Compassion, also referred to as "compassionate love" (Sprecher & Fehr, 2005) and "communal responsiveness" (Clark & Monin, 2006) is a close cousin of sympathy but is at once more encompassing and more enduring, and involves attempting to understand another in need as well as an inclination to help and support the needy other (Sprecher & Fehr, 2005). Compassion could

be to involve the form of love the Greeks considered *agape* and includes related concepts like caregiving, selflessness, and altruism (Berscheid, 2010). Fromm (1956) states that two important requirements of love are respecting and knowing another, and the former cannot occur without the latter. We argue that compassion, then, is the closest theoretical concept to his enduringly popular psychoanalytic take on love, and as a result of being so comprehensive and additionally involving respect and understanding, it is the concept we consider most important for schools to measure accurately for the benefit of internal goal-setting and demonstrating added value to stakeholders.

"Compassionate love" refers to a particular feeling, but compassion itself is best understood as a character trait, and more specifically, as a virtue. The virtue of compassion, when appropriately activated, can give rise to particular episodes of compassionate feelings and, in turn, compassionate actions. But compassion itself is a stable disposition that a person can have (to some degree) even without it being activated (C. Miller, personal communication, August 2013).

Compassion should be understood as a virtue that can give rise to action, and it is theoretically distinct from a term such as "pity." Until recent times these two terms were used interchangeably but pity now refers to a feeling that is much more "shallow and motivationally idle" in the lexicon of modern times (Crisp, 2008). A person who "feels bad" for a homeless person but does nothing to help is more readily considered to be showing pity as opposed to a person who feels this feeling but then acts upon it, who we could claim was showing compassion.

But it is important to note that compassionate feelings without the corresponding virtue (or ideal) of compassion may not necessarily be virtuous, and resulting actions may not be as well. For instance, if a person were to feel compassion for a disabled person, but feels sorry for this person as a result only of how she looks rather than because of the wider burdens she experiences on a day-to-day basis in society, then this person is not feeling compassion in the right way, although he is certainly feeling it at the right time. It could also be argued that if a person sees photographs of dead children from a war zone but only feels a small bit of sadness and concurrent compassion, then he is not feeling the right amount of it to be considered virtuous (Crisp, 2008). Compassion, the enduring virtue, can quickly become something harmful or ineffective if it is not felt and

acted upon in the right ways, for the right reasons and in the right amount (Miller, 2013; Crisp, 2008).

The last term to discuss in relation to compassion is "altruism." Altruistic motivation to help another comes from concern for the good of that other person, regardless of whether the person acting will benefit or not (C. Miller, personal communication, August 2013; Batson 1991). What follows is often, but not necessarily, altruistic action to relieve the other person's distress. But without being coupled with the virtue of compassion, altruistic motivations and resulting actions could be harmful, often unintentionally so. In sum, altruistic motivations and resulting actions that are non-harmful are the products of the virtue of compassion.

Enhancers and Inhibitors:

We will circumscribe our discussion of causes and correlations of compassion to the actions that result from it. We do this because of our unique position as consultants in school settings. When thinking carefully about the outcomes most commonly worked toward with children in school, including showing respect for others, cooperation and collaboration, and service to the community (all interpersonal in nature), we argue that attempting to measure simply *feelings* of compassion toward another will not be adequate. Save for asking someone how they feel in response to someone else, there is little way of empirically measuring feelings anyway, although there is much research on the physical (facial expressions, body language) and physiological (heart rate, skin temperature) reactions that indicate feelings of "other-oriented concern" for distressed others (Fabes et al 1993; Holmgren, Eisenberg & Fabes, 1998; Zahn-Waxler et al 1992). This type of research (usually involving heart rate monitors and electrodes attached to the chest) is very difficult to conduct on a wide scale in a school setting, however. And once again, simply knowing whether or not a student has feelings of concern or compassion is not enough for us without resulting altruistically motivated action, and we argue that schools would agree.

In this way, we will focus here on the factors that lead to helping, or "prosocial" (as opposed to "anti-social") behavior, although we recognize that there is still debate in the research literature as to whether these behaviors can ever be altruistically motivated (Batson & Powell, 2003;

Piliavin & Charng, 1990). These behaviors include helping, comforting, sharing and cooperating. Altruistic prosocial behavior is sometimes viewed as a subset, including self-sacrificial behavior and behavior for which there are no obvious personal rewards (Batson & Powell, 2003).

There is an area of research on the "moral emotions" of guilt and shame that is applicable here. Tangney et al (2007), in their review of the relevant literature, distinguish between guilt, a negative evaluation of a specific behavior, and shame, a negative evaluation of one's sense of self, as two key emotions that affect whether someone responds prosocially. Their review of research shows that, for the most part, guilt motivates one to act prosocially more often and shame reduces this potential outcome. There is also theoretical discussion, but little empirical research, on feelings of disapprobation (negative judgment) toward immoral behavior as a potential motivator (Prinz, 2012).

Temperamental and dispositional variables that affect prosocial responding include sociability, social intelligence and assertiveness (Eisenberg, Fabes & Spinrad, 2006), ability to read others' emotions properly (Knight et al, 1994), moderate but not aversive levels of personal distress (Holmgren, Eisenberg & Fabes, 1998; Eisenberg, Wentzel & Harris, 1998; Batson, 1991), sufficient emotion regulation to handle personal distress (Eisenberg et al, 1999), a positive "others-concept" or "prosocial value orientation" which includes feelings of personal responsibility for others' welfare (Staub, 2005; Grimley & Zucker, 1991), one's complexity of moral reasoning ability (Malti et al, 2009), one's beliefs about social norms and roles, such as a belief in reciprocity (which enhances helping) or one to "mind your own business" (an inhibitor) (Batson & Powell, 2003), a generalized sense of trust for strangers (Yamagishi & Yamagishi, 1994), and level of moral identity (Pratt et al, 2003; Arnold, 1993; Colby & Damon, 1992).

Situational cues that affect levels of prosocial responding include perceiving that one is interdependent with a needy other (Irwin, Mcgrimmon & Simpson, 2008), what other observers are doing nearby (the "bystander effect") (Latane & Nida, 1981), and levels of guilt and shame induced by situations unrelated to the experimental condition in which helping was possible (Miller, 2013a; Prinz, 2012).

Operationalization and Measurement:

Although we have settled on the term "compassion" for our conceptualization of the virtue of love, there is little empirical research using the same term, save for the work by Sprecher (Sprecher & Fehr, 2005). Close theoretical cousins, empathy and sympathy, have been assessed much more. Interestingly, although empathy is consistently defined by researchers as not requiring the sharing of a negative emotional state such as sadness or pain, most empathy measures involve assessing the degree to which another person's suffering evokes helping behavior (see Eisenberg & Miller, 1987, for a review on empathy).

Empathy and sympathy are most commonly assessed by asking participants to report on their levels of these feelings in varying situations. Some ask participants to respond to global statements that tap both cognitive (i.e., perspective-taking) and affective (i.e., tender or caring feelings) components of empathy (childtrends.org, 2013; Lawrence et al, 2004; Bryant, 1982; Davis, 1983). Many of the studies that use these global self-report measures combine them with peer and teacher reports as well (see Eisenberg & Miller, 1987 for a review). Self-reported emotions in response to stories of distressed others have also been used, and empathy has been operationalized as the degree of match between the story protagonist and the participant in the study (Zhou, Valiente & Eisenberg, 2003). Lastly, researchers have used emotionally evocative audio or video segments of distressed others and told participants that the situations are real, not hypothetical, and then asked questions about how a participant was feeling (see Zhou, Valiente & Eisenberg, 2003 for review). There is also at least one personality subscale that assesses one's perceived level of active concern for others' welfare (Costa & McRae, 1992).

Observed behavioral reactions (i.e., facial expressions, vocal responses, gestures) to distressed others have also been used (Zahn-Waxler et al, 1992), but suffer from consistently inaccurate coding by researchers and have most often been used only with young children, so there's less ability to generalize to wider populations.

Years of strong empirical work on physiological responses to distressed others by Eisenberg and her colleagues makes the strong case for these (i.e., heart rate and skin temperature changes) as

additional robust measures of empathy (Zhou, Valiente & Eisenberg, 2003 for a review; Holmgren, Eisenberg & Fabes, 1988).

There are many ways in which helping behaviors that *grow out of* empathy, sympathy and compassion have been assessed throughout the research literature, but technically they aren't measuring the feelings inside someone else but rather the resulting actions that come out of these feelings. Regardless, when paired with other measures described above, they seem quite adequate. This could take the form of signing up to help out a person who recently experienced a personal tragedy (Batson, 1991), sharing with a peer or unknown other (see Eisenberg & Miller, 1987, for a review), donating part or all of one's payment for participating in a study (Sturmer et al, 2006), volunteering after being asked to help task unrelated to the experiment (Eisenberg et al, 1991), or even just observed "spontaneous helping" by researchers in a natural classroom environment (Eisenberg et al, 1999; Eisenberg et al, 1991).

There are also a group of measures that assess self-reported frequency of "prosocial" behavior by either asking about specific behaviors from the past (i.e., "I have donated goods to a charity") (Rushton, Chrisjohn & Fekken, 1981) or by presenting vignettes and asking whether the helping behavior acted out by the protagonist in the story would be something the person would likely do (Shelton & McAdams, 1990).

Finally, there are self-report measures that assess aspects of personality that relate to altruism, kindness and empathy (Ashton et al, 2007; Goldberg et al, 2006; Costa & McRae, 1992).

Part III - Our Proposed Approach

Diligence:

For Students:

The student app we have developed to utilize Ecological Momentary Assessment methodology is provisionally entitled INoted. It allows researchers to send multiple-choice questions with up to 8

answer choices to sub-groups of students. The app sends a global phone alert to the student saying a question is ready to answer. The student simply taps the alert or opens the app to have the question appear with answer choices below it. The app has a built-in system that keeps track of how many questions a student has answered. Once a certain number of questions are answered (this threshold can be set by researchers) a \$10 Amazon.com gift code is automatically sent to the email address on file for the student. There is also a dynamic News Feed page that can be populated with the latest news from the school (See Appendix C-5 for screenshots of all features).

Not only can these multiple choice question alerts ask about current behavior, mood and attitudes, the app can also be used to ask standard self-report questionnaire items used for decades in social science research one at a time rather than at one sitting. We believe this will significantly eliminate fatigue effects that have dogged researchers for as long as survey methodology has been utilized. Another perennial issue for studies relying on survey data collection is low response rates, where follow-up paper or online surveys go unanswered in significant numbers as studies progress. We believe that these apps' ease of use combined with asking only one simple question at a time along with the built-in monetary incentive feature will virtually eliminate low response rates. In fact, in pilot research conducted in the Summer of 2013 with a group of 20 graduating 8th and 9th graders from an independent boys' school, we have consistently received response rates averaging near 90%, which is at least 15% more than in other EMA studies that have used cellphones to collect data (Courvoisier, Eid & Lischetzke, 2012) (See Project Description for full discussion of this pilot work).

We propose to measure diligence and grit amongst students with a variety of questions send to the INoted app about daily behavior and attitudes. To measure grit, which we define as diligence applied to long-term goals, we plan to send an initial series of questions to assess if they have any long-term goals (i.e., "How much of a priority is it for you to get in shape this semester? How much of a priority is it for you to get better grades in History class?") We will then ask regular follow-up questions (one to two times per week) about specific behaviors related to those goals:

If a student claims to want to get in better physical shape, we will follow up once a week for three months with questions like "Did you engage in vigorous physical activity for at least an hour today?" (Yes/No) (See screenshots of individual questions in Appendix C-5). Similarly, for those committed to improving their grades, we plan to ask an initial set of questions to assess a baseline ("What is your current GPA?" or "What is your current overall grade in science class?") but also provide an option of "Ask me tomorrow" each time.

By providing an option of "Ask me tomorrow" we can assess levels of procrastination, which we would argue is an aspect of behavior that inhibits diligence. We can then ask questions about behaviors related to working hard on a subject. An example might be "Do you have a science test coming up this week?" and for those who answer "Yes," then following up with the nightly question of "How long did you spend studying for your Science test tonight?"

As for day-to-day diligence, we plan to ask questions addressing behavior that is necessary to complete work in a diligent manner. For instance, we will mimic work done in previous studies (Trautwein et al, 2009) by asking regular questions such as "How hard did you work on your homework tonight?" and "How careful were you on your homework tonight?" Although this still allows for subjective appraisals of one's experiences, it at least eliminates the need to reflect on a long time period, like most self-report questionnaires. We will also ask such questions about a subject that a student previously reported was his least favorite or most challenging, which we argue will assess level of persistence in the face of difficulties, boredom, and being overwhelmed. We also plan to ask questions about specific behaviors or events from a day or week that point to a lack of care and attention in one's daily activities (i.e., "Did you lose points for a careless mistake on an assignment today?").

In an attempt to delve into the all-important contexts and motivations around which humans exhibit various virtues, we plan to ask a follow-up question to students who stated that they worked hard on a given day as to *why* they did this ("What is the major reason you worked so hard today? (To avoid punishment by parents, To get a great job someday, To avoid looking like a failure to friends and classmates, etc.)).

For Teachers:

The teacher app, entitled SoNoted, similarly pushes 3 to 5 questions per day to teachers based on the students seen that day in classes. Questions can be one of two types. "Type 1" questions ask about one student at a time on our three domains of Diligence, Compassion and Honesty simultaneously (i.e., "How would you rate John Smith on the following three domains today?"). Each domain is rated using a five-point continuum slider that is defaulted to the middle, neutral position. Teachers must then swipe either left or right depending on what they remember about the student's behavior that day. Teachers receive an alert in the morning prior to that day's classes telling them the two to four students they will be rating at the end of the day. In this way, a quick snapshot of that day's behavior is recorded with minimal disruption to a teacher's responsibilities.

SoNoted has been specifically designed to utilize a repetitive and simple question-asking framework in order to minimize the potential for teacher attrition due to being overwhelmed by "one more thing" to add to their already lengthy set of duties to perform each day. This makes it impossible to ask specific questions about the teachers' mood or attitudes to correlate with their daily ratings of student behavior. We recognize this as a limitation but we believe that, combined with student data, these daily "snapshots" will be so much more accurate than typical school progress or grade reports that the app will prove its worth without this contextual data able to be captured.

We will name one of the continuum slider scales "Diligence" and utilize the following descriptive language below each of the five points on the slider:

Very lacking in diligence	Somewhat lacking in diligence	Neither	Somewhat diligent	Very diligent
---------------------------	-------------------------------	---------	-------------------	---------------

(See screenshots of Type 2 question examples in Appendix C-6).

After calculating based on the average student load for middle school teachers in American independent schools (45-60 total students across all class sections), this leads to a student being assessed by a single teacher with a Type 1 question about every 7 to 10 days, or about 5 times in

a typical quarterly marking period. Adjustments to this question-asking algorithm will be made for public school teachers who have larger class lists.

Type 2 questions, alternatively, list a *group* of students and ask a teacher to decide whether any of the students in the list exhibited a specific behavior that day in a "small way" or a "big way." (See screenshot in Appendix C-6). Each of these Type 2 questions will be sent once a day, on a continuous looping cycle, to either a random subset of a teacher's full list of students or to just one class section (i.e. first period History class). This looping cycle will mean that if there are, for instance, five total Type 2 behaviors we want to assess, then the question-asking algorithm will cycle back to the first of the five behaviors on the sixth school day (See Appendix C-7 for sample six day question-asking cycle for one teacher).

We envision this type of question to be used for less frequent and more memorable behaviors related to diligence such as:

- "Carelessly rushed through his/her in-class assignment"
- "Started a project but then gave up without finishing it."
- "Persisted with an in-class task, even after failing"

Measures for Validation:

After sifting through the research literature, we have settled on the following self-report scales against which to validate our smartphone measures of diligence:

1. Grit scale (Duckworth et al, 2007)
2. Persistence and Prudence subscales of VIA-YOUTH (Park & Peterson, 2006)
3. ChildTrends Positive Indicators, Diligence scale (childtrends.org)
4. Behavioral Indicators of Conscientiousness (BIC), Industriousness and Laziness subscales (Jackson et al, 2010)

The Grit scale has been used extensively in public charter school populations (Tough, 2012) and we have further found in our own consulting work with schools over the past two years that the term "grit" is mentioned consistently by most school administrators when discussion turns to character and its assessment. The Persistence Scale for Children, in our view, includes survey items that most closely parallel our conceptual definition of diligence (i.e., "I won't try to solve a problem again and again if I don't find the solution in the first time I try it."), and therefore, we chose it over other similar scales. The BIC burrows down even more specifically into behaviors that a person must reflect upon rather than global statements, and therefore, it would be interesting for us to compare data from this to our daily assessments of students' behaviors through their smartphones. Lastly the ChildTrends Positive Indicators scale, while short (only 5 items), has massive norm data to compare to from an ongoing nation-wide project funded by both federal agencies and private foundations.

Our behavioral measure of diligence is to be determined but will most likely be inspired by work from DeWall et al (2011) who had students persist with difficult multiplication problems for a set time period.

In addition to these established measures, we plan to also compare our Ecological Momentary Assessment measures using smartphones to indicators of diligence that schools have historically kept track of, such as quarterly progress reports that include scale items on perseverance and hard work ("Needs Improvement-Satisfactory-Excellent" is a typical scale used). We also plan to comb narrative grade report comments for indications of diligence on specific assignments (i.e., "John showed great persistence on his Science Fair project over the last month of school, even coming in to school on a Saturday to finish it up!").

Honesty

For Students:

Our main approach for assessing honesty amongst students using Ecological Momentary Assessment will be to simply ask them "Did you lie or stretch the truth today?" and "Did you

cheat on something today (test, quiz, sports contest, etc.)?" The key here will be not so much asking them these questions, but how we follow up.

The research literature on honesty is clear about enhancers and inhibitors of the tendency to lie (see Miller, 2013a for a comprehensive treatment), and we hope to delve into these factors with focused follow-up questions for students who admit that they lied on a given day. We will regularly ask questions about whether or not a student lied that day "in a small way" or a "big way" (to mirror work by DePaulo and Kashy (1998) who distinguished between "everyday lies" and "big lies") but then we will make sure to follow up with questions such as "What was the reason you lied today?" providing answer choices that cover reasons cited in relevant research (Miller, 2013a; DePaulo & Kashy, 1998) (See screenshots in Appendix C-5).

We will also follow up with those students who lied with questions asking about their perceptions of the situation, such as:

- "How bad did you feel telling the lie?"
- "Do you think you were successful in telling the lie?"
- "Did you plan to lie ahead of time or was it on the 'spur of the moment?'"

We will use a similar approach with instances of cheating, so that we follow up with those who admit to cheating with questions that probe for their motivations (i.e., "To avoid feeling like a failure", "To gain a competitive advantage over others", "Because it was thrilling or fun to try not to get caught").

To assess consistency across context, we plan to ask questions about lying and cheating that varies the recipients of these acts as well as situations, such as "Who have you been most dishonest toward this week?" (Family members, close friends, kind-of friends, strangers, I've been honest with everyone equally).

As a result of our definition of honesty being tied up in one's beliefs about themselves as honest, we argue that we may also need to first gather from each participant information about what they

value and how important these things are to their sense of identity. We could then look for whether or not descriptors like "Fair" and "Honest" are important to individual students' sense of self. To do this, we will look into measures of moral identity (Barriga et al, 2001; Arnold, 1993) that ask participants to assess how important key moral concepts are to their sense of self. This could then be used as a covariate in our data analysis.

Our "paper and pencil" battery of tests will also ask students to reflect on their global beliefs about lying and cheating (i.e. "If done for the right reasons, lying and cheating is okay" (Schlenker, 2008) and "Everybody lies, it's no big deal" (Nas, Brugman & Koops, 2008)). We believe this will provide us with data that will help us delve into the contexts in which students are motivated to exhibit these behaviors. There will also be items sprinkled throughout the test battery that assess social desirability bias, which we argue is another form of lying or "stretching the truth" about oneself.

All in all, we believe our Ecological Momentary Assessment protocol for students will directly address Peterson & Seligman's (2004) admonition that "researchers should attempt to assess whether actual behavior is consistent with one's morally justifiable principles" (p. 262). In this case, we will be assessing how important being fair and honest are to an individual student, and then comparing this to self-reports of lying and cheating on a periodic basis with the INoted app.

For Teachers:

Similar to the approach we will use for diligence and grit, we will send teachers daily Type 1 questions on individual students that ask for ratings on continuum slider scales for each virtue, one of which will be "Honesty." (See screenshots in Appendix C-6). The scale wordings will be as follows:

Very dishonest	Somewhat dishonest	Neither	Somewhat honest	Very honest
----------------	--------------------	---------	-----------------	-------------

In addition we will send Type 2 questions about specific behaviors related to honesty where we ask teachers to indicate whether or not specific students from a list exhibited dishonest behaviors

in either a "small way" or a "big way" (i.e., "Did anyone in this list lie or break a promise to you today?") (See Appendix C-6 for screenshots).

Measures for Validation:

We will utilize the following self-report and behavioral measures to validate our smartphone apps, modified for use with middle schoolers by occasionally simplifying language and/or omitting items that only pertain to adults or older adolescents:

1. Items pertaining to lying from the Integrity Scale (Schlenker, 2008) and How I Think Questionnaire (Nas, Brugman & Koops, 2008; Barriga et al, 2001);
2. Authenticity subscale of the VIA-YOUTH (Park & Peterson, 2006);
3. Fairness subscale of the HEXACO-IPIP (Ashton et al, 2007);
4. Academic Integrity Survey for High School (McCabe, 2008);

The Integrity scale and How I Think Questionnaire will provide us data on a student's overall beliefs about how wrong dishonest behavior actually is. Does he believe it's okay in certain situations and not others? This will be key for correlating with our day-to-day assessments of motivations for dishonesty outlined above. We are also glad to see that the Integrity scale attempts to minimize the possibility of social desirability bias by wording items so that none of them seem overtly negative (i.e., "Lying is sometimes necessary to accomplish important, worthwhile goals"). The VIA-YOUTH and HEXACO-IPIP subscales are well-established personality scales with a large amount of comparison data, the former specifically with adolescent populations and the latter with an international, mostly adult dataset. Lastly, because we are in a school setting and will be asking many daily questions about cheating in school contexts, we have included McCabe's measure for academic integrity. We especially like this measure because it not only asks about frequency of past cheating behavior but also about the situations where a student believes cheating is acceptable.

Our behavioral measure of honesty is to be determined but we are intrigued by work done by Mazar, Amir & Ariely (2008) who handed participants both a test sheet and an answer sheet for a

search task in which they are asked to find two numbers in a matrix of nine that add up to 10. They hand in their answer sheet to experimenters but then can take money from an envelope in whatever amount they want for each correct answer, thus giving them a chance to lie and take more money than they are actually supposed to receive. We will need to carefully plan how to do this in a school context, perhaps without use of actual money but with another similar material good of value to students at each school.

We will also supplement our "paper and pencil" and behavioral measures of honesty with any other reporting mechanisms for dishonest behavior at individual schools, such as Honor Council violations, incidents logged with the school counselor or principal, or mention of any lying or cheating behavior in quarterly or semester-wide progress and grade reports.

Compassion

For Students:

We propose to measure compassion by asking daily questions about attitudes and specific behaviors that delve into the "depth and breadth" of compassionate concern and related helping behavior people show to others (Weissbourd, R., personal communication, August 24th, 2013). For breadth, we refer here to how wide a person's circle of concern goes outside his immediate family and friends. It's one thing to ask a student if he helped his friend with homework last night, but would he also be equally willing to help the socially awkward, new student with no friends? Does he feel any compassion for the plight of people in other parts of the nation or world or just the student at the end of the lunch table sitting alone? What about people who are fundamentally different than him in identity or background? (i.e., if he is Jewish, does he feel concern for Palestinians as well as Jews in Israel/Palestine?)

There is also an issue of depth of concern, something Batson and his colleagues (1991; 1989) have illuminated empirically in their research on empathy and altruism and Miller (2013a) has convincingly formed a comprehensive theory around. Simply put, research has consistently shown that there are varying degrees of helping that most people elicit depending on how much

it will cost them. Getting back to our hypothetical student from above, if a friend asks for help on homework, is he willing to spend only five or ten minutes or does he offer to block out his entire afternoon? A simple way to encapsulate this is to ask "Who does one have compassion for, and how much?" (Weissbourd, R., personal communication, August 24th, 2013).

We also recognize the importance of correlating helping behavior with one's moment-to-moment mood or feelings of shame or guilt, based on our reading of the research literature that shows strong connections between these feelings and variations in willingness to act prosocially.

With all of this taken into account, our Ecological Momentary Assessment protocol for compassion will take many forms. We will assess specific behaviors like whether or not one comforted someone he didn't know too well that day or deliberately hurt someone's feelings within the past 24 hours. There are issues here with socially desirable responding, and we plan to sprinkle throughout the protocol questions from established social desirability scales for children to control for this (Paulhus, 1991; Crandall & Crandall, 1965). We also believe that the data collected from teacher reports through the SoNoted app will further help us calibrate the measurement protocol for social desirability.

We will also convert items from scales such as the Self-Report Altruism Scale (Rushton, Chrisjohn & Fekken, 1981) into questions about daily behaviors ("Did you volunteer to let someone go in front of you in line today?" or "Did you spend time volunteering your time to the school community this week?"). We will also ask questions about behaviors that reflect thoughtful, but not required, behavior toward others ("How many messages of thanks (both on paper and through electronic communication) have you sent to people in the past month?")

As for assessing one's breadth of concern, we plan to ask questions about a series of current events from local to global (i.e., How concerned are you about the recent massacre of civilians in Syria?) and combine these responses into a score that reflects that person's ability to have compassion for a variety of people across situations (natural disasters, physical violence, poverty). We will also ask questions about who one reached out to in her immediate school environment (i.e., "Is there a new student in your class this year?" and if "Yes", asking follow-up

questions like "Did you reach out to a new student today to help acclimate them to school?"). We will also ask questions about demographic differences between the respondent and those she helps ("Have you comforted or stood up for someone of differing sexual orientation than you in the past week?" or "Did you give a compliment to someone you didn't know too well today outside your friendship group?").

To assess one's depth of care, we will ask questions like "Did you help a friend on his homework last night?" and if "Yes", then follow-up questions like "How long did you help your friend on his homework last night?" We will also ask questions to illuminate the degree to which an act of helping cost another (i.e., "To what degree was helping your friend a burden to you?").

For Teachers:

Just like with the other two virtues, we will send teachers daily Type 1 questions on individual students that ask for ratings on continuum slider scales for "Compassion" (See screenshots in Appendix C-6). The scale wordings will be as follows:

Very uncompassionate	Somewhat uncompassionate	Neither	Somewhat compassionate	Very compassionate
----------------------	--------------------------	---------	------------------------	--------------------

We will also send Type 2 questions so that teachers can assess a group of students at one time asking "Did any student in this list..." and following up with examples of compassionate actions such as:

"Spontaneously help another student in need today without you prompting?"

"Comfort another student he/she didn't know too well?"

"Share a valued possession with another student who seemed to need it?"

Measures for Validation:

We plan to validate our smartphone-based measures of compassion against well-established self-report measures that relate closely to the domain:

1. ChildTrends Positive Indicators Empathy scale (childtrends.org)
2. Compassionate Love Scale (Sprecher & Fehr, 2005)
3. Strengths & Difficulties Questionnaire (SDQ), Prosocial Behavior subscale (Goodman, 2001)
4. IPIP-NEO Altruism subscale (Goldberg et al, 2006)
5. Visions of Morality scale (Shelton & McAdams, 1990)
6. Self-Report Altruism scale (Rushton, Chrisjohn & Fekken, 1981)

The ChildTrends four-item empathy scale, while short and potentially unable to measure all aspects of empathy, has vast amounts of norm data on North American adolescent populations with which to compare to. There is fairly limited research on compassion proper, and we are therefore limited to one promising self-report scale from Sprecher and Fehr (2005), the Compassionate Love scale, with two 21-item subscales for "close others" and "all of humanity." The SDQ is similar to the ChildTrends scale in that it has a well-established track record in the psychological literature and also focuses more on self-reports related to helping behavior. The IPIP-NEO is similar in content to the SDQ but is a more typical personality trait measure with international norm data from a wide population.

We also will validate our EMA measures of compassionate helping behavior against retrospective self-report measures of the same, although we will modify each of these for use with middle schoolers in a school setting by simplifying language and eliminating items that describe situations relevant only to older children (i.e., items referring to having an after school job or driving a car). For this purpose we will utilize the Self-Report Altruism Scale (Rushton, Chrisjohn & Fekken, 1991) and the Visions of Morality scale (VoM) (Shelton & McAdams, 1990).

These scales, combined together, provide a comprehensive approach that covers frequency of past behavior, assumptions about one's hypothetical behavior in imagined scenarios, and reflections on one's personality characteristics related to helping behavior. Furthermore, the

Strengths and Difficulties Questionnaire has a version specifically designed for follow-up use, which will be helpful to our proposed study design. One important final note: the Visions of Morality scale includes some items that we would argue assess integrity more than compassion in that they describe situations in which a student "stood up for what is right." An example: "I read where a large company's policies have victimized the lower class of an overseas country. This company makes one of my favorite snack foods. As a way of protest, I give up eating this snack." We chose to eliminate these types of items, but all in all, we see the scale as a promising measure in that it assess private, interpersonal and "social" morality, which mirrors the Compassionate Love scale's emphasis on "close others" and "all of humanity." In assessing these three forms of helping, we also argue that it mirrors our assessment strategy outlined above of delving into both "depth and breadth" of compassion.

We plan to assess students with two behavioral measures of compassion, one that assesses compassionate feelings and another to assess altruistic giving behavior. We are drawn to research that assesses reactions to distressing situations, often operationalized in psychology experiments by showing film clips designed to induce feelings of concern for protagonists (Holmgren, Eisenberg, & Fabes, 1998; Stifter et al, 2008). We plan to look into how to feasibly do this in a group setting in each school and will most likely slightly modify the measures from the relevant research literature to do this. We also want to include a behavioral measure of altruistic helping, and the Ring Measure of Social Values (Liebrand & McClintock, 1988) and Prosocial Value Orientation game (Van Lange et al, 1999) seem most able to be modified for use in group settings (See Appendix B for full descriptions of our preliminary ideas).

Finally, as with the other two virtues of diligence and honesty, we will utilize school reports of compassionate acts, whether from awards students win for "good deeds" or attendance lists from volunteering events or simply from combing through quarterly progress and grade reports that are already part of the established school routine of assessment.