UNIVERSITY OF BIRMINGHAM

Measuring Virtue Better
A Little Later and a Little Rougher

Stephan Ellenwood

These are unpublished conference papers for the 'Can Virtue Be Measured?, held by the Jubilee Centre for Character and Values at Oriel College, Thursday 9th – Saturday 11th January 2014. These papers are works in progress and should not be cited without author's prior permission.

School of Education
University of Birmingham
Edgbaston
Birmingham

Measuring Virtue Better,
A Little Later and a Little Rougher

Stephan Ellenwood
School of Education
Boston University

January, 2014

Abstract for Measuring Virtue Better, A Little Later and Little Rougher

Prior to developing sound ways to measure virtue fine distinctions must be made about both virtue and measuring. The several intricate and overlapping responsibilities of schools, families, and communities in developing virtuous understandings and behaviors in the young make measuring the impact of particular programs a complex challenge. In addition to carefully reviewing both traditional and expanded definitions of measurement procedures we must insure that educators are allowed to design creative and innovative programs prior to developing measuring plans. Two oft-neglected measurement approaches, longitudinal and qualitative studies, afford rich opportunities for determining the effectiveness of wider range of curriculum reforms than is customarily considered in program assessments. Too often the convenience and lure of short-range assessments attract educators and policy makers to curriculum projects that may not have an enduring impact in the lives of the students. Important advances have been made in strengthening the validity and reliability of qualitative, longitudinal measurements. They may not have the exact precision of short-term, quantitative measuring devices, but our understanding of complex matters such as understanding virtue and virtuous behavior can be greatly enriched with qualitative, longitudinal study.

Introduction

In his 1954 essay "Knowledge and Wisdom" Bertrand Russell described the essence of wisdom as "emancipation, as far as possible, from the tyranny of the here and now." (Russell.) We have just reached the 60th anniversary of that advice which, given current infatuations with haste, would qualify it as an ancient text. In drawing his distinctions between knowledge and wisdom Russell provides important guidance for educators' endeavoring to determine how to measure something complex and truly non-hasty such as virtue. Pragmatic educators who are trying to expand, enrich, and deepen what young people study in schools are grateful for Russell's "as far as possible" qualifier. This gratitude is particularly true for those attempting to develop specific comprehensive curriculum reforms dedicated to students knowing, understanding, and becoming virtuous. Such efforts have been undertaken in the face of massive traditional customs long in place for both curriculum and measuring learning.

Education reformers devoted to such substantial change have been at it for quite some time and come from several diverse perspectives. Lately, those reformers have come to realize that although the diversity of perspectives and terminologies has value and must be respected, that diversity at the same time often enables both the massive status quo as well as those more actively resistant to successfully parry reform efforts with challenges along the lines of "Some speak of virtues, others of morals and ethics, still others of character; what exactly is it that you want to accomplish?"

Reformers respond with increasingly improving answers to that important question, but they cannot match the exactitude found in more traditional curriculum provinces such as mathematics or reading. Measuring systems there are deeply rooted, but even if critics come to be mildly satisfied with the answers about the exact goals of moral education, the next, and usually immediate, challenge arrives as "How will you ever measure that?" The question carries a clear implication that it if cannot be measured, then in a data-driven era we shouldn't teach it.

These basic questions about what it is to be accomplished and how it will be measured require large answers that live outside the narrow "knowledge" in Russell's "here and now." In Peter Shaffer's *Lettice and Lovage*, Lettice constantly comically disdains what she sees as "the mere." She is actually affording wisdom to the reformer's best responses to these two basic questions. In addition to unifying diverse reform efforts these reformers would be wise to work with the very best; that is, a Gold Standard for the diverse curriculum efforts to teach virtue, ethics, morals or character and a uniform Gold Standard for measuring impact with a sophisticated, qualitative-longitudinal design.

A review of some of the major reform efforts allows identification of commonalities, opportunities, and problems in curriculum over the past half century. The broadest over-arching term that has emerged from reform programs is character education. Clear insights into ethics, morals, and virtues are indispensable to good character education. But character education includes something beyond ethics, morals, and virtues. High character

involves making good judgements which itself is a level or two beyond skills at decision-making. High character additionally includes Aristotle's familiar "habits of right action." And it involves sensitivity and perceptivity in relating with others as well as an ability and inclination to anticipate long-range consequences of judgment and actions. Thus, character education includes more careful attention to recent developments in social-emotional learning than the more traditional terms - virtue, morals, and ethical learning. Those approaches customarily focus more on the academic, analytic, and abstract.

We cannot turn attention to longer-range thinking and making good judgements without confronting the influence of a rapidly accelerating technology in our global culture. The digital-electronic era clearly affords many opportunities for becoming more richly informed with data and ideas. But it carries its own hidden curriculum – a tendency toward haste and hecticness. Max Frisch described technology as a way of "arranging the world so we don't have to experience it." The finest character education, on the other hand, requires getting as close as possible to reflective learning from meaningful experience. Sound moral judgments depend on thoughtful reflection, taking time, and active communication about nuances, connotations, understandings, and consequences. In his novella *Slowness* Milan Kundera warns us that "speed is a form of ecstasy the technical revolution has bestowed up man." (Kundera, p. 2) Whether it is a false ecstasy or not is a difficult issue, but it differs markedly from the rich ecstasy of understanding and becoming dignified individuals of good character and sound judgement.

Across the last half-century significantly different forms of ethical-moral-virtue-character education have developed. And they have each been subjected to different kinds of assessment and evaluation. They each still have many adherents, but the degree of dedication on the part of individual schools and teachers varies greatly from year-to-year as a consequence of a variety of changing pressures. For instance, the current clamor for STEM (i.e., acronym for Science, Technology, Engineering and Math), which actually could, but rarely does, include productive opportunities for character education, has had a negative impact on the advances many have made in moral-character education. So often a crisply measurable education reform such as STEM steps in line ahead of a more risky emphasis on character education, in part, because the latter requires more complicated measuring.

Mark Twain meets at least two standards for an intellectual: he is amusing and he is clarifying. With respect to character education he is amusing and partially clarifying. In *Following the Equator,* he observed that "It is noble to be good; it is still nobler to teach others to be good – and less trouble"(Twain, 1989, p.1). For far too many the instruction of virtues appears easy. Those who have pursued it carefully, especially in schools know that it is extremely complex. But Twain was fundamentally correct; it is noble, and we have found ways to help students achieve some moral nobility.

The Historic Baseline

In the early days of mass public education those who saw the task as easy reigned. To them moral education took one of two general forms.

1. *The value-neutral tradition* can be summed up best in the simple declaration, "In a free society schools cannot teach values." In this tradition, schools devoted considerable effort to making clear to students that formal education and individual values are separate. Formal education was seen to be the domain of the intellect. Values were believed to derive from more subjective and elusive matters of emotions, ethnicity, religion, culture, and family. Teachers became reluctant to assert any values, in part because a free society supposedly meant no one should ever impinge in any way upon someone else's value system. The belief persisted that schools could meet all of their responsibilities by teaching only what is verifiably true. Thus the separation of the intellect from the student's character development was deeply engraved.

2. *The school as inculcator of basic values* tradition held that schools must teach the importance of living by basic codes that include fundamental and seemingly straightforward, virtues such as honesty, fairness, kindness, and courage. Usually, these were not part of the explicit classroom curriculum, but were the basis of the school's discipline/behavior code. Schools conducted explanations of virtues sporadically and unevenly. Seldom were these virtues accorded the priority of classroom academic instruction. Nor were the teachers very often involved in leading students to understand the complexities and subtleties of a particular virtue. And even less frequently did they help students sort out the delicate choices between competing good choices or choosing the lesser of two evils.

The confusion about teaching values in schools is reflected in recent, frequent discussions with young liberal arts college graduates studying to become teachers. They were asked to rank the priority of important educational goals for the public schools. This group had declared their intention to become teachers and clearly understood the educational system where they had been so successful. Among commonly cited goals for schools, students were most hesitant about setting "to teach morals" as a high priority. They usually ranked it last or next to last. However, another group of students with the same background and professional aspiration using an only slightly modified list of the same goals ranked the phrase "to teach about morals" as one of the most important goals of schools. In the discussions of this prioritizing it was clear that the first group was deeply afraid that in teaching morals, schools meant indoctrinate. To the second group, the responsibility "to teach about morals" drew strong support because then teachers could discuss moral issues without compelling students to agree. If talented young people successful in schooling and eager to become teachers equate the verb "teach" with "instill," "inculcate," and "indoctrinate," great care must be taken. Education reformers need to explicate both this common misunderstanding of the verb "teach" and the simplistic dichotomy that schools can only either indoctrinate or avoid values.

In a free and diverse society, ultimately it is actually not possible for schools to mandate morals for students. But neither is it possible to remain purely value-neutral. Any institution or community operates on a set of values and priorities that it ineluctably communicates in a variety of ways to members of that community. Schools are no exception. No matter what schools, do, even if they remain totally silent on moral or value issues, they communicate a set of values to young people. Fortunately, there are many options beyond teaching vaguely about morals, ignoring morals, or indoctrinating morals.

Three Reforms: Values Clarification, Cognitive Moral Development and Literature Based Character Education

Over the last four decades three broad movements have emerged in the domain of ethics-values-moral-character education. These are in addition to those who persisted in advocating for the tradition of instilling basic, universal virtues. Each of these movements were evaluated and researched with uneven and varying degrees of success. No one has yet produced the ideal curriculum in the domain of moral-character education. The best option for discerning the impact of each reform curriculum would be to measure each as it is conducted optimally with a uniform, richly designed assessment scheme that respects the complexity of good character and the importance of it lasting over considerable time.

One major challenge to the values-neutral and the inculcating values traditions was the values clarification movement. To the originators of this movement, the curriculum was not concerned "with the content of people's values, but with the process of valuing" (Simon et al., 1973, p. 19). The values clarification movement has received considerable criticism, but it did make significant contributions in restoring the importance of values in schooling. It represented the first major challenge to the belief that schools could teach in a value-free manner. Advocates of this reform claimed it would be wiser for teachers to discuss directly social and personal values, but they did not address ways in which careful examination of values could be integrated into the curriculum.

This theoretical construct is based on a three-stage process so students learn how to understand values in a personal or public problem. In the first stage, the choosing stage, teachers encourage students to select freely from alternative choices after consideration of the consequences of each alternative. In the second stage students publicly announce their conclusions. The reformers claim, with some validity, that prior to this program the schools customarily went to great lengths to avoid any acknowledgement of the role of values in decisions students make and therefore have made a grave error by severing schooling from the rest of life. In doing something about their choices, students are participating in stage three, which they label as acting. Teachers must not only help students analyze an issue, but also help design a course of action mindful of the impact on others. Their model further calls for making these actions part of a "life pattern." In doing so they emphasize that decisions should not be a casual collection of independent choices, but that each decision must be situated in a thoughtful system of choices about basic values and actions.

As with many education reforms, the differences between the plan and the actual practices in the field became considerable. Though some of the values clarification reform improves upon past practices, three basic challenges emerged. First, nearly all of their vexing incidents for discussion are slim, ahistorical, and artificial. Second, the specific teacher strategies are very uneven in quality and mostly result in the teacher withdrawing from the discussion. Third, the discussions are oriented too heavily toward personal authentication and too open-ended. Too often students develop a basic impression that their choice really does not matter because all options can be explained. The incidents do not have the texture of real-life science, mathematics, literature, biography, or history, so the thinking engendered is often thin. Real-life decisions are complex, textured, evolving, and involving. Too often, the values clarification situations are simplified, artificial, and sterile.

Harvard's Professor Lawrence Kohlberg took values education in a completely different direction. He repudiated values-clarification as too relativistic. His plan relied heavily on students' stages of development he saw as sequential. At each stage, students conduct a completely different type of thinking than was available to them in an earlier developmental stage. In the first stage, the obedience stage, students determine whether an act is right or wrong strictly in terms of what the teacher or parent permits or forbids. During the second stage, the students come to grips with interpreting rules so they can think about the spirit and basic premises of the rules, in addition to the letter of the law. In the third stage students determine the justice of an action in accordance with what the rules permit or forbid, what the person's intentions were, and the larger social moral values of justice that may or may not be represented in the law.

Kohlberg's work became powerfully influential in the larger values education movement. In many schools, his plan replaced the values-clarification model. He predicated his reform on the analytic processes necessary for students to resolve moral dilemmas. The ultimate goal is universal justice, which is seen as independent of cultural differences. In the Kohlberg plan, the implications for teachers are complex. In the reality of students in classrooms, the lines between the stages are blurred almost beyond recognition. But the basic belief that students proceed through them sequentially is worth some consideration. Significantly, research on this moral cognition model has suggested that students with extensive peer-group participation have moved more productively through the stages than those without such group work.

The strengths and weaknesses of this cognitive moral development approach are epitomized in a moral dilemma often used by teachers. In simplified form, a young German woman in Nazi Germany is faced with the opportunity to possibly save her young Jewish friend. To do so, she will have to break German law and risk her fate, as well as the fate of her family, and her friend. In determining a course of action, the Kohlberg-trained teacher is urged to provide a rationale and a clear, concise set of classroom procedures that reveal to students the strengths and weaknesses of each option. These procedures include rational/analytic questions and some general contextual questions. However, the overall strategy omits any sustained discussion of how this larger European crisis came to be and,

more specifically how it could have been avoided. Further, though students are urged to come to closure and set their own individual response to the challenge – "What would you do?" – the model does not necessarily recommend that students publicly state their choice. This classroom practice could set a pattern of overly individualized and thus less civic approaches to public concerns. To many, it reveals the extent to which value decisions for Kohlberg are more analytic exercises, rather than real, intricate, and humane judgments demanding a consistent respect for subtlety and complex interplay of emotions, conflicting values, interpersonal relations, and historic factors.

The heavy emphasis on discrete cases can lead to an ahistorical and decontextualized habit of mind in students and, eventually in the public mind. It reminds us of the Harvard Law School Dean who claimed that the persistent study of discrete cases in law schools is like trying to educate horticulturalists by only allowing them to study cut flowers. This curriculum overly emphasizes rational, logical, analytic thinking in academic cases while omitting what one would actually do in a complex, real, contextualized situation.

One of Kohlberg's Harvard colleagues, Carol Gilligan, has presented a different, important, and very specific challenge to his moral development model. She claims that Kohlberg completely omits the "morality of caring" that characterizes women's approaches to the kinds of dilemmas found in Kohlberg's materials. She asserts that Kohlberg's base of empirical data, upon which his curriculum is built, derived from a developmental study of 84 boys over a period of 20 years. The result, she concludes, is that the very traits that have defined the "goodness" of women; that is, their care and sensitivity to the needs of others would "mark them as deficient in a highly analytic model of moral development." Kohlberg's putatively higher and better stages, Gilligan claims, are inadequate to the lives of women in which their moral problems arise "from conflicting responsibilities rather than from competing rights and requires for its resolution a mode of thinking that is contextual and narrative rather than formal and abstract" (Gilligan, 1982, p. 19).

The Kohlberg reform frequently disparaged moral education models that were indoctrinating, that were relativistic, or that were merely a "bag full of virtues." However, many have begun to reassert that a set of core virtues is a valid form of character education – as long as the education is rigorous, thoughtful, and respectful. These reformers recognize that there is a form of moral judgement that includes more than rational analytic procedures. It involves a sense of integrity or judgment, what judges call "justice with mercy," a kind of humane integrity that includes and extends beyond rationality, beyond individual feelings, and beyond a check list of individual virtues. To the familiar traditional individual values they have added attention to a sense of social justice and caring. A literature-based character education curriculum realizes that there are subtle, but significant, differences in the value system each culture holds and practices. In one important sense, core virtues are central to the mission of schools because the continuation of the culture depends upon it. That is a comparatively easy pedagogical achievement in a monocultural society. But a stable multicultural society depends upon the successful interaction of many subcultures that value diversity, progress, and growth. Studying

literature from the perspective of character development provides rich opportunities for understanding such complexity and opportunity.

The literature-based character education model recognizes the importance of subtlety and nuance in both real-life interactions and in relationships between characters in literature. The most effective vehicles for an enriched understanding of these experiences are often fiction and biography. When students experience the complex course of actions and choices characters make in a believable work of literature, they gradually come to understand the important and intricate interplay of rational, humane, caring, courageous, and cultural factors in judgments and decisions.

In his 1986 book, *Actual Minds, Possible Worlds* (Bruner, 1986) Jerome Bruner describes two basic ways people make sense of experience and construct reality. Bruner's two modes of mental functioning are propositional thinking and narrative thinking. Propositional thinking accords closely with what we usually mean when we discuss cognitive functioning. It is the kind of thinking that schools encourage most frequently in classrooms. This cause-and-effect thinking which Bruner describes as a "logico-scientific" attempt to arrive at conclusions which are abstract and context-independent.

Narrative thinking is enmeshed with people and events, with time and place. It is concrete and context-dependent. To think narratively is to think in story form. Actions and ideas are lived out in the intuitions, intentions, decisions, and experiences of each individual. While propositional thought may be more highly regarded for many human ends and in academic settings, narrative thinking, in many ways, is more fitting and more effective in developing complicated moral understandings in the lives of young students.

Michael Novak (1970) further explained the importance of literature and culture in influencing our moral compasses. He restores the importance of individual exemplars in education.

> Contemporary studies in ethics, especially in Anglo-American philosophical circles, concentrate upon logic and language. I wish, instead, to concentrate upon the drive to understand and upon the myth of symbols. My reason for doing so is that men seldom, if ever, act according to principles and rules stated in words and logically arranged. They act, rather, according to models, metaphors, stories, and myths. Their action is imitative rather than rule-abiding. Prior to their intention to obey sets of rules, they are trying to become a certain type of person. (Novak, 1970, p.26).

The insights of Bruner and Novak provide important new perspectives for current and future work in character education. Previous programs such as value-neutral, indoctrinating, values clarification, and cognitive moral development have each communicated directly or indirectly that moral, ethical, and values decisions are either independent of, or dependent upon, the analytic intellect. Literature and biography can provide the productive connections between affective and cognitive ways of thinking.

Good teachers have met their responsibilities when their students realize through careful consideration of rich stories that analytic thought can inform intuitions and feelings, while at the same time intuitions and feelings can enrich logical analysis.

The most effective use of literature occurs when stories are vivid, vexing, and vicarious. By "vivid" we mean that the reader actively cares about the characters in a story. If we involuntarily think about a story the next day, then it is vivid. It is the difference between entertainment and art. For adolescents, particularly, the story is vexing when it discloses a problem to be solved, when the main character must grow or decide not to change or grow. If students read about believable, memorable characters in stories and biographies, they are less likely to succumb to the conclusion that the course of their life is dominated by large, uncontrollable forces. Instead, they will see how it is usually influenced more by small daily judgments best grounded in a moral pattern, which often leads to Novak's "becoming a certain type of person." They also come to learn that they are each writing their own story. They are making choices that have consequences.

Literature provides the means by which students can see how believable, complex, interesting people think about values in difficult situations. Literature engages the teachers so that students see them as adults who can add an experienced perspective to such matters. This is a departure from other approaches in which the teacher is either neutral, didactic, or merely a "devil's advocate." Thus, in this model, teachers are enabled to fulfill one of their most important obligations of teaching, serving as an exemplar of clear thinking and responsible action. Doing so requires logical analysis, humane judgment, a sense of nuance and detail, an understanding of the discipline involved in holding to principles under pressure, and experience judging between long range and short range consequences of actions. Those who ask students only what they think or only how they feel about a problem are providing incomplete instruction.

The literature–based model of character education often highlights various dimensions of basic virtues such as honesty, courage, kindness, fairness, caring, and social justice. Diverse literary selections reveal the complexity of each virtue. For example, images of kindness can range from gentile manners to simple altruism, to generous altruism, all the way over to darker forms of manipulative kindness that Charles Dickens described as "rapacious benevolence." Thus, a teacher's injunction to students that they should be kind is indeed a good idea, but not sufficient to the complex thinking students must conduct about countless nuanced choices that come before them.

Not only must students see that each virtue is complex and usually embraces a spectrum of meanings, but also they soon come to realize that good judgment involves knowing how to decide between competing values. Students do not have difficulty deciding that honesty is preferable to dishonesty. Any of the reform efforts under consideration here will be largely successful in inclining the students toward honorable values and away from dishonorable ones. However, when the standards of honesty, for instance, conflict with the standards of kindness in a particular situation, students need help in sorting through the nuances and implications. For example, in simple but intensely personal situations, such as asking children how they like Aunt Betty's pie while in Aunt Betty's

presence, we are asking for subtle judgments to be made quickly. Simple injunctions to students that they should be honest and they should be kind are useless. Rich stories of such seemingly simple incidents would tell us we need to assess the kind of person Aunt Betty is, the child's basic relationship with her, and much more. They give students vicarious experiences in making good judgments.

Judgment is more than problem-solving skills and decision-making skills. Good judgment often seems abstract and ineffable to students. They understand that it has to do with real, but somewhat remote, concepts like integrity, wisdom, and experience. In Harper Lee's classic, *To Kill a Mockingbird*, much of this is made compelling and clear. As students see Atticus Finch bring his sense of courage, humanity, respect, and justice to bear in an urgent situation, they realize that high principles are worth holding, thinking about, and putting in to action. Each character is richly drawn. Teachers who help students reflect on how various characters in this story acted as they did are greatly enriching their students' character development.

Alan Gurganus elegantly explains literature's importance as a way of understanding: "Literature tells us that imagining each other is not just a luxurious feat achieved by rare artists, but a daily necessity for all of us." (Gurganus, 1999, p. 20). His imagery of us "imagining each other" every day captures the foundation of understanding, caring, and behaving well when we are involved with others. He is in fact clarifying how the finest character education programs embrace analytic intelligence, narrative intelligence and social-emotional intelligence. Judge Learned Hand once observed "the spirit of liberty is the spirit that is not so sure it is right" (Hand, 1953, p. 190). Being not so sure we are "right" generates the wondering and attention to detail that enlightens the nuances and complexities of wise judgments. Abstractions and generalizations are often too vague and inaccessible to help students learn about the complex interplay of thoughts, emotion, and language. In good stories this interplay is part of the behavior of believable characters in believable circumstances. Starting there, students can more easily grasp how to transpose subtle ideas to their own lives. Good teachers stop at various points in stories or films to consider what will each character will do next, what should each character do next, and eventually what each student would, as well as, should do next. They must go on to have their students include the reasoning and evidence behind as well as the wisdom of each answer. These issues are visited again at the end of the story so students can evaluate why their estimates were confirmed or not.

Measuring in a Qualitative, Longitudinal Design

Before turning attention to the specifics in qualitative and longitudinal research two considerations are important. The first is considering the implications of these two approaches for measuring different versions of virtue-moral-ethical-character education. Several responsible diverse programs have evolved over the last four decades. To understand the broad field more productively we would be best advised to study at least one site at which each kind of program has been operating optimally for some time. The term "operating optimally" cannot be defined perfectly, but to be identified as a Gold

Standard program it should be integrated in to the regular curriculum and the climate of the school over a period of time, the teachers should have received some preparation in how to integrate the principles of the specific program in to their classrooms, the program should reach across at least four of the grade levels in the community's K-12 program, and the school leaders should have declared a continuing commitment to both the curriculum to a thorough qualitative-longitudinal analysis of the program's impact. If we can thoroughly measure the impact of several diverse optimally-operating programs, then we are better able to identify strengths of various approaches and look in to the wisdom of gradually strengthening and possibly combining them.

The second consideration before turning to the specifics of qualitative and longitudinal design involves looking into the interior of two key terms employed by educators: measuring and accountability. A careful review of each discloses important connotations that extend and enrich the narrower definitions researchers, educational leaders, and the general public commonly intend. To most educators trying to measure impact their governing images and connotations turn immediately, and usually exclusively, to the kinds of measuring scientists, mathematicians, and business people do searching for the specific impact of a specific intervention. All these reside in the quantitative research domain and seek a specific causal relationship between a precise intervention and a specific outcome. Unfortunately, important images and connotations involved in another usage of the term measure –as in "take the measure of a person or situation" - are seldom evoked. Generally educators would benefit from this larger meaning of the term measure. For the diverse programs in the complex field of character education the benefits would be enormous. Educators who consider this broader meaning could begin to think about enlarging both what is taught and what is learned. Of course, they would quickly discover that "taking the measure" of a broader educational goal such as good character in a student is imperfect compared to mastery of a narrower concept or skill. But as educators learn so often, the perfect is the enemy of the good. Is it really more valuable to perfectly measure something quite modest than to "take the measure" of something important and complex even if the measure is rougher?

Similarly, the term accountability frequently motivates educators to be sure their work can demonstrate incontrovertible effectiveness of a particular program. Often accountability morphs into a matter archly stated as "showing that the taxpayers are getting their money's worth." If the interventions and the learner's mastery are defined narrowly, then the likelihood of showing a high-frequency impact, and therefore, a narrow accountability is high. But, as Einstein regularly cautioned, "Everything that can be counted does not necessarily count; and everything that counts cannot necessarily be counted." A larger version of accountability derives from common parlance as in "give an account of…" This suggests a narrative-centered explanation where the phenomenon being explained is important and complex. Causation may not always be perfectly clear as in simpler, more narrowly bounded behaviors, but we cannot divert educators away from understanding complex phenomena simply because we connote accounting so narrowly. By making the activities of educators that reductionist we risk that their interactions with learners will become too heavily reductionist. The noted research scholar David Berliner has warned

persistently that the pressure for testing narrows the curriculum and makes it more superficial**.** Good character education simply cannot be narrow or superficial.

Two excellent opportunities exist for educators to make their assessments of curriculum interventions more significant. They involve accepting assessments that are little rougher than the current heavily quantitative studies and that are conducted over a longer period of time. There are likely very few basic educational goals more suitable to well-designed qualitative, longitudinal assessment than character education.

Both qualitative and longitudinal studies have been around for quite some time. As we review their applicability to character education programs we also need to consider suspending a common custom in professional circles. It is usually bad professional manners, while asserting a new and less familiar point of view to also criticize and challenge the traditional models. More typically advocates advocate while completely refraining from criticizing other practices. Such restraint may have negative effects. It suggests that traditional measuring and assessment remain just as they have been and eventually become a contest between the two approaches – quantitative and short range on one hand and qualitative and long-range on the other hand. Such a contest would deprive each from learning from the other. For instance, perhaps learnings traditionally assessed quantitatively such as mastery of quadratic equations could also be assessed qualitatively and over a longer period of time. It is entirely possible that a set of students could show mastery of that particular concept one semester and have no capacity for it two, four, or six years later. Correspondingly, a well-taught character education program might show only modest progress at the end of a semester, but could show frequent usage and growth over two, four, or six years. We know what happens when teachers and students focus on curriculum content with the understanding that the most important demonstration of student learning is a short-time away – the Friday quiz or the monthly test. Imagine how the teaching and learning would be designed and unfold if in all courses the most important demonstration of student learning would be two, four and six years after graduation.

The mantras of data-driven decisions are incanted often in education circles these days. Rarely do we hear mention of principle-driven decisions. But if we can enlarge the definitions of data to include qualitative information and long range impact, then we can move the education profession more in the direction of rich, textured, understandings and deeper explanations of complex teaching and learning. In doing so we will have opened better opportunities for complex assessments that will improve character education and perhaps widen the lens for current assessment practices across the entire curriculum, so that teachers teach and students learn with an eye to deeper understanding and longer range impact.

Over the past 25 years qualitative research has come of age in education. It emerged as a response to heavily quantitative measures that were limited to students demonstrating whether they had gained mastery over pre-determined and highly precise concepts and skills. Experienced educators have long known that in good lively classrooms students learn much more than specific, predetermined concepts. Qualitative assessing

developed ways for evaluator-researchers to uncover deeper, broader, and more naturalistic learning. Whatever qualitative measuring, or more accurately assessing, lost in objectivity and precision it gained in richer understandings of what students did learn. The narrowness of pure quantitative measuring too often became a search for what the students don't know. In general, limiting measuring to the quantitative paradigm misses much of the subtlety of diverse ancillary learning that frequently is influential. For a character education curriculum that emphasizes the importance of complexity and nuances the opportunities for rich understanding afforded by a careful qualitative study, conducted longitudinally, must be thoroughly limned and considered. In doing so we must review the conventional wisdom which holds that objective measures have accuracy, independence, and veracity subjective measures can never attain. Examples of rich qualitative data came from students answering open-ended questions about healthy caring relationships in a literature-based character education curriculum. Two responses illustrate powerful ancillary learning. One student wrote: "I'm not sure what I want to be when I grow up, but I don't want to be like Judy." (a character in a story used in the curriculum.) another wrote: "I learned that relationships are more complicated than computers." This curriculum is centered on *The Art of Loving Well, A Character Education Curriculum for Today's Teenagers.* It includes over 40 literary selections and classroom activities that focus on helping students develop a wide range of healthy relationships.

In their valuable challenge to high stakes testing, *Collateral Damage*, Sharon Nichols and David Berliner, point to a study of students who took the renowned TIMSS (Third International Mathematics and Science Study) exams (p. 186-87). Interviewers found that 58% of the students understood more than what was on the exam and would have answered more questions correctly if they were in conversation and could ask about the problem. On 29% of the correct answers students conversely did not really understand why their answer was correct. On only 13% of the items were interviewers convinced that the item disclosed what the students knew. At the very least this should caution us about assuming that quantitative measures are rigorous while qualitative measures are somehow soft.

Qualitative research has emerged largely from anthropology, investigative journalism, and clinical psychology. In these fields, when conducted optimally, these researchers rely on a variety of instruments to construct a broad, deep, and rich account of complex evolving phenomena. In assessing the longitudinal impact of any of the character education programs described previously researchers could include some of the customs in quantitative research. For example, a control group that did not use the curriculum could be identified and tracked. Also a form of the pre-and post-test model easily could be set up. Good qualitative research is time-consuming, but there is not any good way to get sound, reliable, valid, and meaningful information about the impact of a complex curriculum except to examine it from several angles. One of the angles could well be an information/knowledge-centered paper-pencil examination given before the teaching, at the conclusion of the teaching, as well as two, four, and six years after the teaching. A thorough plan could also have participants complete an open-ended short answer questionnaire which could in turn could be part of a face-to-face interview. In an optimal assessment this quantifiable information would be integrated with what evaluators learn

through interviews at each stage and student journals across the full 10-year time span that includes a four year instructional program. If we use the literature-based curriculum as an example, then researchers could evaluate what students recall over the several years of the longitudinal study to determine if one story, one character, or one problem became a sort of talisman for students. They could also assess the habits of mind that the curriculum generated by asking participants to read a new short story to identify the degree to which the participants could transfer any understandings or recollections from the original direct teaching and apply it to a new situation. If a significant number of participants demonstrated the ability to transfer insights to new situations and the ability to do so over a period of time would have to establish any program as successful. Similarly, skilled interviewers could discuss with participants how they have handled moral dilemmas at two, four, and six years after the original teaching. Unprompted information from participants is of more value than unprompted information. But qualitative researchers can gain insight after prompting participants to determine if any of the impact is somewhat latent. Asking if participants recall any character from any story is one kind of information. Identifying a character and talking over any possible meaning that character has for the student is also valuable information.

In terms of understanding the impact or the effect of any education program the Gold Standard measurement scheme also would have to include studying the long range influence on the lives of the learners. It means that long after the period of direct teaching the learner would demonstrate a rich understanding of the key ideas, the ability and inclination to transfer those key ideas to new situations, good judgment about consideration of alternative courses of action, and the courage and skills to take the right action. Across the culture and in the education profession we have heard a great deal about Big Data to help understand and predict sweeping phenomena. Only very recently have we heard as much about Long Data. But the New York Times on December 8, 2013 (Business, p. 3) reported on the powerful insights gained from tracking data over time. Character education reformers would do well to turn the attention of the entire profession toward the impact of programs across students' later lives, not just their lives in school.

One obvious and important downside to such a plan is that it is expensive and does not accord well with contemporary culture's widespread lust for instant results. And it carries the risk of multiple intervening variables impacting a treatment group over several years. Thus, the standards for declaring a project as impactful must be modest. On the positive side if a longitudinal study is carefully designed and implemented it can provide information so we can truly understand about the ultimate goal of education programs; that is, that they influence the lives of the learners long after schooling.

There are several approaches to longitudinal studies, but for the purposes of the richest understanding of character education programs the best option is what qualitative researchers identify as panel studies. This means that researchers examine the same study participants over as long a period of time as possible. It requires a large pool to account for a likely evaporation rate that can undermine a study. The longitudinal research design must be qualitative in that it involves an interaction and adjustments for the researcher and for the participants as time unfolds. This design can productively involve comparisons

with a control group; that is, students with no focused character education program. Good longitudinal qualitative studies include paper-pencil surveys, interviews about the survey and about the participant's daily life as well as the participant's reflective journals over time, participant's responses to triggered recollections and classroom portfolios. If the curriculum and the assessment plans are well constructed and well-supported, and if it includes diverse assessment mechanisms, educators can point to either a high quality program that has a lasting impact or we will find out how to improve the intervention program and/or the complex measuring system to improve results. But in this design we will be working in the area of complex education goals, programs, and measuring rather than keeping each component very simple to honor a declared need for objective data. In the fields of character education objective data about longitudinal impact is almost impossible to produce. Critics of longitudinal studies say that the time delays allow for nearly uncountable intervening variables affecting outcomes. But if the results are affirming it does not mean that the original program was without impact. Qualitative and longitudinal studies more commonly come up with correlations rather than direct causations. The mantras about "Correlation is not causation" echo. But it is an oversimplification. Edward Tufte sagely points out that to be purely accurate the statement would read "Empirically observed covariation is a necessary but not sufficient condition for causality." If that makes one yearn for Standard English Tufte generously translates, "Correlation is not causation, but it sure is a hint." (Tufte, p. 45) A hint of causation and the opportunity to assess the magnitude of the hint should encourage those curriculum providers and measurers of programs who want to venture into new complex, consequential territory. It can illustrate the ways in which the achievement of a complex good triumphs over the need for perfect measurement.

Among many well-regarded, influential longitudinal studies is the US Framingham Heart Study. Like many medical longitudinal studies this began as a descriptive study originally relying on 5200 healthy adults. Since 1948 this continuing study now reaches across three generations. This moved from descriptive and correlational to causal study as the epidemiology of cardiovascular disease was reconstructed around the discovery of risk factors such as blood pressure, smoking, cholesterol, obesity, diet, and limited exercise. This, of course, is more than one could hope for, or perhaps even need, in studying good character. But it affirms the power and importance of going well beyond Lettice's "mere." It is a rich qualitative, longitudinal study that has unquestionably saved lives. Our challenge is to come up with the curricular moral equivalent.

References

Beyer, B. (1976). Conducting moral discussions in the classroom. *Social Education,* 40 (4), April 1976.

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.

Ellenwood, S., McLaren, N., Goldman, R., Ryan, K. (1993). *The Art of Loving Well, A Character Education Curriculum for Today's Teenagers.* Boston: Boston University (School of Education.)

Gilligan, C. (1982). *In a different voice.* Cambridge, MA: Harvard University Press.

Gurganus, A. (1999). Our golden sacrificial couples. *Harper's.* February 1999.

Hand, L. (1953). *The spirit of liberty: Papers and addresses*. Irving Dillard, ED. New York: Alfred Knopf.

Kundera, M (1996). *Slowness* (L, Asher, Trans.). New York: HarperCollins Publishers, Inc. (Original work published in 1995

New York Times, December 8, 2013

Nichols, S. and Berliner, Dr (2007). *Collateral Damage: How High Stakes Testing Corrupts America's Schools.* Cambridge, MA: Harvard Education Press

Novak, M. (1970). *Experience and nothingness*. New York: Harper and Row

Russell, B. "Knowledge and Wisdom", (1954). www.personal.edu.kent.edu.

Shaffer, P. (1987). *Lettice and Lovage*. New York: Harper and Row.

Simon, S., Howe, L., Kirschenbaum, H. (1972). *Values Clarification.* New York: Hart Publishing Company.

Tufte, E. (2006). *The Cognitive Style of Power Point: Pitching Out Corrupts Within.* Cheshire, Connecticut: Graphics Press

Twain, M. (1989). *Following the equator: A journey around the world.* Hartford, CT: Dover Publishers, Inc.