



**Aligning Assessments in Character Education:
Integrating Outcomes, Implementation Strategies,
and Assessment**

Marvin Berkowitz

These are unpublished conference papers for the 'Can Virtue Be Measured?', held by the Jubilee Centre for Character and Values at Oriel College, Thursday 9th – Saturday 11th January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**

Aligning Assessments in Character Education:
Integrating Outcomes, Implementation Strategies, and Assessment

Marvin W. Berkowitz

Sanford N. McDonnell Professor of Character Education

Co-Director, Center for Character and Citizenship

University of Missouri-St. Louis (USA)

Presented at

“Can Virtue Be Measured?” Conference

University of Oxford

January 2014

Character education is a growth industry around the globe. More and more countries, local education agencies, schools, and educators are interested in intentionally focusing on how best to foster the positive development of youth, along with promoting their academic learning and growth. While the specific term “character education” may not always be the rubric under which such endeavors are placed (c.f., Berkowitz & Bier, in press), the project remains the same. Unfortunately, as the aphorism “the road to hell is paved with good intentions” is sometimes understood to suggest, just because educators are embracing character education as a critical aspect of effective schooling does not mean it is having the desired effects on student learning and flourishing.

In this paper, some of the core challenges in moving from a sincere intention to foster student flourishing (virtue, character, values) in schools to actually succeeding in doing so will be addressed. A specific focus will be on how the assessment of virtue/character/values is critical to bridging the intent-to-impact gap that is so common. Included will be considerations of: (1) what counts as evidence of effective practice; (2) the challenge to generate and/or access empirical knowledge to guide practice; (3) the importance of fidelity of implementation; (4) the frequent failure to align implementation strategies with outcome goals; (5) the various targets of assessment: implementation; climate; outcomes; (6) the challenge to optimally operationalize the assessment targets. One important reason for coupling a consideration of assessment with a consideration of implementation threats is that they inform each other. Often when asked about how to design an evaluation of a character education program, the first answer is that it needs to be embedded in a logic model of outcome goals, implementation strategies, and the justifications linking them. Only then can evaluation be optimally designed.

How do we know what to implement?

This conference, and hence this paper, focuses on assessment. Hence, it is a bit beyond the scope of both to delve deeply into how practitioners (or other decision-makers of implementation)

decide what to implement in order to foster the development of virtuous students of character. In fact it is a bit circular too, because assessment is critical to practical decisions about implementation strategies. In a sense, we are breaking unnaturally into a circle here. Good theory and research are needed to craft and select effective practices. But those practices have to then be tested with appropriate research to insure that they remain effective when transplanted to a new educational context. So practice and assessment rely upon each other.

This is precisely why we proposed and executed the “What Works in Character Education” (WWCE; funded by the John Templeton Foundation) project from 2000 to 2004. We wanted to answer the question of what research could tell us about effective practice in character education. As the WWCE project has been well documented (e.g., Berkowitz & Bier, 2005), it will not be revisited here. Subsequently, others have also examined this question, and we have recently reviewed the information for a forthcoming book chapter (Berkowitz & Bier, in press). From these various sources, we have identified 19 groups of practices that research supports as effective character education implementation strategies (see Table 1). Hence, those wishing to craft or select an implementation strategy could turn to such scientific sources to aid in the selection.

The Problem of Technology Transfer

One of the challenges, however, is that the substantial body of research on implementation effectiveness is only minimally impacting practice. One of the classic examples of this is in the field of drug prevention, a field that is often related to or part of character education initiatives (Battistich, 2008). In the US (and the UK), one of the primary drug prevention programs is called Drug Abuse Resistance Education (D.A.R.E.), which provides local police officers, trained in the D.A.R.E. program, to deliver a drug prevention curriculum in schools. Schools quickly embraced this as an effective drug prevention strategy, in part because it also builds school-law enforcement relations and provides free

curricula. The problem is that research demonstrated repeatedly that D.A.R.E. was an ineffective program, and that there were many other programs (including character education programs like the Child Development Project/Caring School Community, www.devstu.org) that are effective in reducing student drug use. In fact, D.A.R.E. was found to be so ineffective that ultimately the US Surgeon General listed D.A.R.E. as an “ineffective primary prevention program” in 2001, and it was later even included in a list of treatments with the potential to cause harm (Lilienfeld, 2007). What is interesting about this timeline is that D.A.R.E. was founded in 1984 and the research reports questioning its effectiveness began to surface in 1992. It is worth noting that D.A.R.E. UK began in 1995, three years after the first report of ineffectiveness surfaced in the US.

When data questioning D.A.R.E.’s effectiveness as a drug prevention strategy were presented to educators, they often reacted with anger challenging the data and affirming their deep faith in the effectiveness of the program. As one educator opined to the author on such an occasion, “I don’t care what the data say. I can make up my own data to support it!” Clearly, getting data to consumers is not all that is needed. You can lead practitioners to data, but you can’t make them think!

Various strategies can support effective character education technology transfer; i.e., adoption of research-based implementation strategies by educators. By simplifying the information, without distorting it, one can make it more palatable and less intimidating. The PRIME model (Berkowitz, 2009) is an attempt to distill best practices down to five core principles (Prioritization, Relationships, Intrinsic motivation, Modeling, Empowerment). The dissemination of the information more effectively can increase the likelihood that practitioners will even be aware of it. Knowing *about* is not sufficient however. Without high quality and on-going professional development to support the research-based implementation strategies, implementation is much less likely to be effective. Modeling by valued others (e.g., school leaders) also helps motivate implementation. In particular, observations of and

messaging from practitioners who have successfully implemented and reaped the desired developmental and/or academic outcomes appears to be particularly compelling to practitioners. For example, staff site visits to exemplary schools is a powerful strategy. In part, this is why the Character Education Partnership (www.character.org) began and continues the National Schools of Character program (NSOC); i.e., to provide examples of exemplary practice with corresponding results. It is also why I take the annual cohort of school leaders in the Leadership Academy in Character Education (Berkowitz, 2011; Berkowitz, Pelster & Johnston, 2012) on site visits to NSOC schools. Finally, empowering educational stakeholders, and most especially those who will administer and implement the strategies, to share in the consideration, design, and selection of effective strategies helps with adoption in general and with fidelity of adoption.

The Problem of Fidelity

When one is prescribed a medical treatment that includes instructions on implementation (e.g, frequency of dosage, conditions of dosage, duration of treatment, sequence of treatment elements, etc.) and doesn't comply fully, it is unfair to conclude that the treatment was ineffective when the symptoms do not improve. Often antibiotics, for example, come with instructions to take a pill a certain number of times per day and to continue until all the pills are gone even if the symptoms disappear. If one were to stop the treatment after a few days because the symptoms eased or disappeared, and then the symptoms returned, again it would be illogical to conclude that the antibiotics failed. Rather, the antibiotics were not implemented with full fidelity so no fair test of full implementation occurred. This happens in education all the time. Educators are trained in how to implement with fidelity, but then only deliver some of the lessons, or do each one partially, or do not use the proper pedagogical techniques in delivering the lessons, etc. If the curriculum then fails to lead to the desired outcomes, such as character development, it cannot be concluded that the intervention, *as designed and*

prescribed, was ineffective; only that the intervention, *as incompletely and/or inaccurately delivered*, was ineffective. We will return to this issue of fidelity in the assessment section to follow.

The Problem of Alignment

Many years ago a high school principal whom I had mentored in learning character education approached me enthusiastically and informed me that his school had begun a bold new step in their quest to foster student character development. The step was the comprehensive commitment to and adoption of service learning as a core pedagogical strategy. Before explaining how, he wanted to tell me why. There had been a cheating scandal in the school and in reflecting upon and researching academic dishonesty, the staff discovered that it is a national (and beyond) epidemic in secondary and post-secondary schools. They realized that they likely had had their heads in the sand, and were quite chagrined to realize this blindness on their part. So they turned to service learning. The principal proceeded to again enthusiastically tell me the deep way they had committed to and embraced service learning as a school-wide platform. They studied it. They invested in high quality professional development. They were mandating it. It was to be embedded in all curricular areas.

After lauding him for the excellent choice of an effective academic method and character education strategy, as well as the wise way in which it was being supported and implemented, I said to him, "But I have one question. What does service learning have to do with academic integrity?" He was dumb-founded. They had leaped from a platform of a serious "presenting problem" (the school cheating scandal) or outcome goal (increasing academic integrity and reducing cheating) onto a landing place of a research-based implementation strategy (service learning); however, they had not stopped to consider the logic connecting the two. They had no logic model. Subsequently, there was no alignment between outcome goal and method (implementation strategy). Service learning is a well-tested method

that has both character and academic outcomes (Billig, 2000; Billig, Jesse, & Grimley, 2008), but it has not been linked to academic integrity.

It is quite typical that schools and districts that commit to character education begin by identifying a set of virtues, values, or character traits as the nexus for their efforts. The first principle of the Character Education Partnership’s “Eleven Principles of Effective Character Education” (www.character.org) is to have a set of consensually chosen core ethical principles, which aligns with CEP’s definition of character as “understanding, caring about, and acting upon core ethical values.” Likewise, CHARACTERplus (www.characterplus.org), a model regional character education organization in St. Louis Missouri (USA), suggests schools and districts begin with a community-wide process of choosing such core virtues. Other character education programs like CharacterCounts! (www.charactercounts.org) provide a set of universal or quasi-universal character traits.

What is interesting is that sometimes these schools align their practices with their outcomes, but often, like the high school described above, they do not. And when they do, it is typically at the most obvious level of delivering lessons *about* their core concepts; e.g., if responsibility is a core concept, they write classroom lessons about responsibility. Rarely, however, do they systematically articulate a logic model connecting their methods with their targeted outcomes. And almost never do their assessments, when they actually engage in assessment, align with their methods or their targeted outcomes. Most distressingly, perhaps, is the fact they have trouble seeing this omission, and, if they care to rectify it, they struggle with creating a reasonable logic model of the connections between implementation strategies, outcomes, and assessments.

Aligning Assessment with Implementation and Targeted Outcomes

At this point it should be clear that the usefulness (and design) of the assessment of a character education initiative depends in large part on the conceptual work that precedes it. Having a clear logic

model is critical to this conceptual work. Typically it is best approached from the perspective of what has been called “backwards design” (Wiggins & McTighe, 2006). While this is not a perfect fit, it helps reframe the issue.

The starting point (the back end) is the outcomes. Defining and justifying the outcomes clearly is where the assessment project and the overall character education initiative begins. Whether one selects the outcomes locally or adopts a set of outcomes from an existing model or example, the argument remains the same. One cannot design, implement or assess without knowing what the character education outcomes are intended to be; i.e., what the initiative is intended to cause. If, as for the high school example described above, the key outcome is an increase in academic integrity, then both the implementation strategies adopted and the assessments utilized must be aligned with academic integrity. Service learning, as a method, is not so aligned as there is no evidence nor theory to suggest that service learning will increase academic integrity. McCabe (McCabe, Trevino, & Butterfield, 2010) on the other hand has empirically identified implementation strategies that increase academic integrity. And unless the school assesses indices that validly reflect on student academic integrity, there is no way to know if the initiative has been effective. Knowing the outcome(s) is the critical first step.

Once the outcomes are identified, then research can be done to identify what methods have been shown to cause those specific outcomes. Only then can an adequate assessment plan be designed. This differs slightly from the curricular design notion of backwards design, which places assessment between outcome and implementation. This is largely because that model focuses mostly on assessment of outcomes. This paper will argue for a more complex and differentiated approach to assessment, including student outcomes but going beyond that as well.

The assessment model proposed here is to assess three separate but related domains: implementation fidelity; moderator variables (especially, school and classroom climate); student

outcomes. Every year in the Leadership Academy in Character Education (LACE), approximately 30 school leaders are asked to end the year of professional development by crafting an assessment plan for their existing or proposed character education initiatives. The plan is required to include the assessment of these three domains (implementation, climate, outcome). Nearly all struggle significantly with differentiating them from each other, with defining the constructs, and with designing the implementation. These struggles occur despite clear and detailed definitions and instructions and the caution that this thought experiment will be challenging (as it is each year with each new cohort in LACE). So, each of the domains will be treated in turn here with a consideration of the common stumbles that educators make in attempting to grapple with each one.

Assessing Implementation Fidelity

Content. Of course, what implementation content is assessed will depend on the specific implementation and its component pedagogical elements. However, one can identify some generic categories of implementation.

- Whether or not the specific implementation category is actually being implemented should be assessed. A few decades ago, during the values education era in US education, a district-wide values education program in a large urban school district was assessed for implementation. The assessment strategy was to survey the school principals and ask them if the teachers in their respective schools were delivering the lessons. The answer was strongly affirmative; the principals indicated that the program was being fully implemented by the teachers in their schools. Unfortunately, when teachers were later surveyed directly, they indicated that they did not like the program and were not implementing it. Apparently the principals were unaware of this and assumed it was being implemented. The measure of implementation (and the source of the data) needs to be one that can actually assess implementation. Asking those principals

about classroom practice was a waste of time as they apparently did not know what was being implemented in the classrooms in their schools.

- Relatedly, the frequency of implementation, where relevant, is important to assess. Many initiatives require or suggest a certain frequency of implementation. For example, daily morning meetings are part of Responsive Classroom. Weekly community meetings are part of the Just Community School model. Just because a certain frequency is prescribed does not mean it is actually happening.
- The completeness of implementation may also be relevant and critical to assess. Some programs have a specific curriculum with a set of lessons. Are all the lessons being delivered? By all relevant educators? Are all components of a lesson being included all the time? For example, a critical part of the experiential learning cycle is reflection (Developmental Studies Center, 1997); however, educators often omit the reflection at the end of the experiential learning lesson/unit.
- The quality of implementation, while a more complex concept, it also important for assessment. Are the implementers doing so in the correct manner, with the correct pedagogical frame, and with the requisite implementation skills? I have witnessed numerous educators using an empowerment strategy like class meetings in a manner that is manipulative and attempts to direct the students to a particular choice or outcome.

Process. How one assesses implementation also will vary widely depending on the nature of the implementation as well as the resources available for implementation assessment. Ideally, implementation will be directly observed, by trained observers. For example, random sampling of classrooms which are visited by trained observers is a good strategy. Alternatively, teachers could log their implementation; e.g., keep a log of when lessons are delivered. This could include who delivered the lesson, what specifically was delivered, which students were present, etc.

Assessing School Climate

Content. It is assumed here that effective character education implementation will lead to changes in the social climate of the classroom and the school, which principally includes the quality of relationships but much more (e.g., shared norms, sense of safety, etc.). Evaluations of the Child Development Project revealed that climate was the mediator between implementation and student outcomes (Schaps, Battistich, & Solomon, 1997; Solomon, Battistich, Watson, Schaps, & Lewis, 2000). That is, students benefited from the Child Development Project implementation if it first led them to experience the school more positively. Hence assessing climate as a mediator may be very helpful in assessing the overall initiative and its impact. The National School Climate Center (NSCC; www.schoolclimate.org) identifies four major dimensions of school climate: safety; relationships; teaching and learning; the external environment. Benninga et al. (2003) found that perception of a “clean and secure physical environment” was one of the dimensions of school climate and character education that best predicted academic success. A comprehensive review of different definitions of school climate is well beyond the scope of this paper, however. Suffice it to say that they are many and varied.

Process. Currently there are many options for assessing school climate, especially in the character education field. It is beyond the scope of this paper to review them, but it is suggested that the reader go to the website of the National School Climate Center (www.schoolclimate.org) for both reviews of the relevant literature and models of climate assessment. These and most others are typically triangulated self-report assessments (from students, educators, and parents). The NSCC, for example, offers the Comprehensive School Climate Survey (CSCI) which has scales for students, parents and school staff to complete. Alternatively, one can use trained observers to assess a school climate.

Student Outcomes

Content. What one assesses as student outcomes should come directly from one's logic model as described above. In the backward design model, the first step is to define what the targeted outcomes are. These then should drive implementation and the assessment of student outcomes. If for example, one adopts the CharacterCounts! Model (www.charactercounts.org), then one should assess, at the minimum, development of the Six Pillars of Character which are the foundation of that model (trustworthiness, respect, responsibility, fairness, caring, citizenship). Certainly other outcomes can be added, as they often are. However, as noted, most character education initiative either do not have an explicit set of outcomes, or do not treat them as measurable outcomes. They become part of implementation, even just window dressing, sometimes literally; but they are not expected to be measurable outcome variable. Instead proxy variables are used as outcomes, and often they are pretty far afield from the character concepts that undergird the implementation model.

The most commonly assessed proxy variables in local assessments of character education initiatives are student attendance at school, student academic achievement and related behaviors (e.g., school drop-out, high school graduation rate), and student discipline data (e.g., referrals to the office for misbehavior, more serious suspensions or expulsions, etc.). While the latter can be closely linked to character, they rarely are. For example, it is rare to see a school disaggregate office referrals by the type of infraction and then link those categories, when relevant and possible, to specific character outcomes in the logic model or general school approach to character education. In scientific research, Berkowitz and Bier (2005), in the What Works in Character Education review, found that the ten most commonly assessed outcome variables were: socio-moral cognition; pro-social behaviors and attitudes; problem-solving skills; drug use; violence/aggression; school behavior; knowledge/attitudes about risk; emotional competency; academic achievement; attachment to school. Yet the most common character traits targeted in schools (and plastered on their walls) are: respect; responsibility; caring; fairness; honesty. It is difficult to directly map the commonly assessed outcomes onto these targeted outcome goals.

Process. This is not only challenging, but at the heart of the rationale for this conference. Ultimately this is akin to measuring personality, something which social scientists have done for over a century, and something we all try to do as human beings. Impression formation and other forms of evaluating others (e.g., do we think the boy our daughter is dating or the new neighbor our son hangs out with is a good influence, a good person, has character). In the US, national teacher certification standards now push for the evaluation of “dispositions.” Certain professions (such as the American Bar Association for lawyers) have evaluations of morality or character. There is an entire industry of personnel assessment in the corporate world around honesty and other character traits. Yet, we are far from having good assessments of character or virtue.

Objective measures abound, and many have strong psychometric properties. They are consistent over time, relate in theoretically predictable ways to related measures, discriminate between groups in theoretically predictable ways, and predict expected behavior patterns. But they are far from ideal. Clinical assessments may be more robust, but are very unwieldy (e.g., require extensive experience with the person to be evaluated, need to be done one person at a time, etc.), require a lot of training, and may vary widely depending on who is doing the clinical assessment.

How one assesses outcomes will also vary depending on the reason for doing so. Criminal justice may want to assess whether a particular incarcerated predator has been reformed sufficiently to reenter society. Schools may want to know if their character education initiative is improving student character. These are very different goals requiring different levels of evidence, and hence different strategies for assessment. Ultimately, the question raised by this conference is indeed an important one. How can we best assess the development of virtue or character?

Conclusion

Ultimate there is no one-size-fits-all prescription for character education assessment. It does require a clear sense of what one is trying to accomplish (outcomes) and how one is going about trying to make that happen (implementation). But depending on the answers to those questions and the resources at one's disposal, the ultimate assessment plan will vary widely. Over 15 years ago, I wrote a Primer on evaluating a character education initiative (Berkowitz, 1998), which is essentially a series of questions any prospective evaluator should ask him or herself before venturing forth on such a project. The gist remains the same; namely, think hard about why you want to do an evaluation and what you want to know before you commit to and then design an assessment plan.

References

Benninga, J.S., Berkowitz, M.W., Kuehn, P., & Smith, K. (2003). The relationship of character education implementation and academic achievement in elementary schools. *Journal of Research in Character Education, 1(1)*, 19-32.

Berkowitz, M.W. (1998). A primer for evaluating a character education initiative. Washington, D.C.: Character Education Partnership.

Berkowitz, M.W. (2009). Teaching in your PRIME. In D. Streight (Ed.), *Good things to do: Expert suggestions for fostering goodness in kids* (pp. 9-14). Portland: Council for Spiritual and Ethical Education Publications.

Berkowitz, M.W. (2011). Leading schools of character. In A.M. Blankstein & P.D. Houston (Eds.), *The soul of educational leadership series. Volume 9: Leadership for Social Justice and Democracy in Our Schools* (pp. 93-121). Thousand Oaks, CA: Corwin.

Berkowitz, M.W., & Bier, M.C. (2005). *What works in character education: A research-driven guide for educators*. Washington DC: Character Education Partnership.

Berkowitz, M.W., & Bier, M.C. (In press). Research-based Fundamentals of the Effective Promotion of Character Development in Schools. In L. Nucci, D. Narvaez, & T. Krettenauer (Eds.), *Handbook on Moral and Character Education*. New York: Routledge.

Berkowitz, M.W., Pelster, K., & Johnston, A. (2012). Leading in the Middle: A Tale of Prosocial Education Reform in Two Principals and Two Middle Schools. In Brown, P., Corrigan, M. & Higgins-D'Alessandro, A. (Eds.), *The handbook of prosocial education: Volume 2* (pp. 619-626). Lanham MD: Rowman & Littlefield.

Billig, S.H. (2000). Research on K-12 school-based service-learning: The evidence builds. *Phi Delta Kappan*, May, 658-664.

Billig, S.H., Jesse, D., & Grimley, M. (2008). Using service-learning to promote character education in a large urban district. *Journal of Research in Character Education*, 6(1), 21-34.

Developmental Studies Center (1997). *Blueprints for a collaborative classroom*. Oakland CA: Developmental Studies Center.

Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53-70.

McCabe, D.L., Trevino, L.K., & Butterfield, K.D. (2010). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11, 219-232.

Schaps, E., Battistich, V. & Solomon, D. (1997). School as a caring community: A key to character education. In A. Molnar (Ed.), *The Construction of Children's Character, Part II, 96th Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.

Solomon, D., Battistich, V., Watson, M., Schaps, E., & Lewis, C. (2000). A six-district study of educational change: Direct and mediated effects of the child development project. *Social Psychology of Education*, 4(1), 3-51

Wiggins. G. & McTighe, J. (2006). *Understanding by Design*. Pearson.

Table 1

Research -Based Practices

		Reviews:	WWCE	WWCE+	Lovat	CASEL	EPPi	Hattie	Marzano
		Outcome:	Char/Acad	Character	Values	SEL/Academic	Academic	Academic	Academic
		Input:	Character	Character	Values	SEL	Citizenship	Mixed	Academic
METHODS									
PRIORITIZATION	Professional Development		√	√				√	
	Common Language				√				
	School-wide Implementation		√	√	√			√	√
RELATIONSHIPS	Interactive pedagogy		√	√	√	√	√	√	√
	Family/Community Partic.		√	√	√				√
	Building Trust			√					
	Focus on relationships								√
	Nurturance			√					√
INTRINSIC MOTIVATION	Real/Relevant curriculum						√		√
	Service to others		√	√	√	√			
	Progressive Behavior Mgmt.		√	√					√
MODELING	Modeling		√	√	√				
EMPOWERMENT	Empowerment			√					√
OTHER	Direct Instruction		√	√	√	√			
	Inquiry-based education							√	√
	Recognition								√
	High expectations			√				√	√
	Safe Schools								√
	Reflection				√				√

WWCE: Berkowitz & Bier (2005a)

WWCE+: Berkowitz (2011)

LOVAT: Lovat et al. (2009)

CASEL: Durlak et al.(2011)

EPPi: EPPi (2005)

HATTIE: Hattie (2009)

MARZANO: Marzano (2003a, 2003b, 2007)

From Berkowitz and Bier (in press).