Research Paper

Tackling the Wicked Problem of Measuring What Matters: Framing the Questions

Yong Zhao, Michael Wehmeyer, James Basham and David Hansen
University of Kansas

Abstract

Purpose: Making policy makers, researcher, education leaders, and assessment developers aware that what matters in education assessment is a wicked problem that cannot be easily solved following traditional approaches.

Design/Approach/Methods: Starting from the questions that what matters in education assessment, this article presented such questions as a wicked problem because there is no consensus, not right or wrong answer, and certain solutions may lead to side effects on students and society. Therefore, a new approach of ecology should be involved, and different education outcomes or intended qualities of learners are presented in complex relationships.

Findings: Deciding what matters in education assessment is a wicked question. It is not a tame or technology problem and can be resolved by any conventional approaches. What is pivotal now is to decipher what matters in education and then what should be measured and ultimately how to measure. The ecology and collaborate approach deliberated in this article could expedite such a process.

Originality/Value: This article advocates paradigm change in understanding and resolving one of the most urgent problems in education. It provides an ecology explanation of the relationships that exist among the different education outcomes and students’ qualities. By guiding through the

Corresponding author:
Yong Zhao, School of Education, University of Kansas, 419 Joseph J. Pearson Hall, Lawrence, KS 66049, USA.
Email: yongzhao@ku.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
dissecting of the problem step by step, this article has demonstrated a unique angle of understanding the wicked problem.

**Keywords**
Assessment, ecosystem, wicked problem

Date received: 11 March 2019; accepted: 6 September 2019

What is measured in education represents a society’s view of what is important for schools to teach and what matters for children to learn. What gets measured reflects the outcomes a society expects of its education system and what its future citizens should be equipped to do. Paradoxically, once an educational outcome is measured, it becomes what matters, even if it turns out to be an unimportant, or irrelevant, outcome. What we measure is what schools and students pursue. As a result, what is measured has a significant impact on the curriculum, the educational experiences of children, and the qualities of future citizens. For example, the *washback effects* of testing on learning, teaching, and curriculum has long been observed as a common phenomenon in language teaching (Cheng & Watanabe, 2004; Shohamy, Donitsa-Schmidt, & Ferman, 1996).

We are on the brink of a significant shift in education measurement that can have long-lasting impact. Recent years saw increasing discontent with traditional measures of content or cognitive skills across a narrow spectrum of subjects (e.g., math and literacy) because these measures seemingly reflect only a minimal part of what is needed for success in the new world (Brunello & Schlotter, 2010; Duckworth & Yeager, 2015; Levin, 2012; Wagner, 2012; Zhao, 2016a). Accompanying this discontent has been a growing awareness of the role that nonacademic (in the traditional sense) capabilities have in student success, resulting concomitant proposals outlining what ought to be measured and taught in schools (Brunello & Schlotter, 2010; Duckworth & Yeager, 2015; European Communities, 2006; Levin, 2012; Partnership for 21st Century Skills, 2007; Wagner, 2008, 2012; Zhao, 2016a). It also appears that policy makers and education leaders are increasingly willing to consider expanding the definition of educational outcomes and, consequently, what should be measured (Perkins-Gough, 2013). Thus, there are signs that the education measurement enterprise will undergo major systemic change and, by extension, so will students’ education experiences.

The list of what ought to be measured is long and wide-ranging. In addition to traditionally measured subjects, such as math, language arts, social studies, and science, there have been calls to add assessments of other subjects, such as arts, music, foreign languages, and financial, digital, or global literacy (Zhao, 2016a). Furthermore, it has been suggested that we also measure other capabilities, variously labeled as noncognitive skills, dispositions, human qualities, or 21st-
These capabilities include a set of sometimes overlapping and potentially competing concepts, including innovation skills (Wagner, 2008; Wagner, 2012), creativity (Beghetto, 2017; Kaufman & Beghetto, 2009), entrepreneurial skills (Aspen Youth Entrepreneurship Strategy Group, 2008; Zhao, 2012), happiness (Seligman, 2006, 2011), physical well-being, self-determination (Ryan & Deci, 2000), social–emotional well-being (Wentzel, 1991), mind-set (Dweck, 2006), grit (Duckworth, Peterson, Matthews, & Kelly, 2007), resilience, communication skills, and collaboration skills (Partnership for 21st Century Skills, 2007; Trilling & Fadel, 2009).

Given the potential gravity of the consequences of expanding educational outcomes, we must carefully define the what in “measuring what matters” before rushing to mandate that schools adopt new measures. Without a clear definition, no matter which technical methods or approaches are developed, efforts to improve education assessment are unlikely to bring about desired outcomes of improving education for all children. But, defining what matters is a wicked problem (Australian Public Service Commission, 2012; Conklin, 2006; Rittel & Webber, 1973).

What matters as a wicked problem

The term wicked problem was first used by University of California, Berkeley design professors Horst Willhelm Jakob Rittel and Melvin. M. Webber in 1973 to describe social planning problems that cannot be successfully resolved with conventional linear analytic approaches (Roberts, 2000). A wicked problem does not mean it is evil. Rather, it means the problem is highly resistant to resolution and any given solution may lead to other problems. In contrast to wicked problems are tame problems, meaning that problems that can be technically complex but can be clearly defined and resolved. The term wicked problems has been widely used to describe complex policy problems such as global climate change, health care, and social injustice.

The first step to resolve or manage a wicked problem is to recognize it as such. Successfully tackling wicked problems requires new ways of thinking about problems and solutions. In this article, we discuss some of the wickedness of the problem of defining what matters in education assessment. Our purpose is to bring awareness to policy makers, researchers, education leaders, and assessment developers, so that we recognize this as a wicked problem that cannot be easily resolved following traditional approaches, in large part because measurement is often treated as tame technical problem that can be solved following traditional linear, analytic approaches. We also suggest a set of questions that can guide collaborative efforts to support better solutions to this wicked problem.

No consensus, not right or wrong answers

One of the defining characteristics of wicked problems is that their solutions are not completely or verifiably right or wrong, but rather better or worse depending on perspective of the particular
stakeholder (Conklin, 2006; Rittel & Webber, 1973). Different stakeholders will have differing views of and solutions to wicked problems, and often there is no clear consensus.

What matters in education is a classic wicked problem. There seems to be consensus that we need to measure what matters, but there is no consensus on what exactly does matter. Disputes over educational outcomes are not a stranger to education. The academic traditionalists have battled the progressive educationalists over academic skills versus child development for over a century (Hirsch, 2010; Norris, 2004; Ravitch, 2001; Wraga, 2001). Different stakeholders have made various proposals: some want to hold onto the foundational skills in core subjects (e.g., math, language arts), some want to add noncognitive skills but still keep the traditional, some want to replace the traditional with the new, and still others want to expand the list of what has been traditionally measured. Each group of stakeholders has their evidence and rationale, and no one proposal can be shown to be entirely wrong or right scientifically.

Even among newly proposed capabilities and qualities (e.g., noncognitive), there exists a wide range of competing ideas, with each proposed list vying to be most important. There is the definition of College and Career Readiness (CCR), a generally accepted goal for today’s schools in the United States, which includes five categories of outcomes: (a) academic knowledge; (b) critical thinking/problem-solving; (c) social and emotional learning, collaboration, and/or communication; (d) grit/resilience/perseverance; and (e) citizenship and/or community involvement (Mishkind, 2014). There are the popular 4Cs—creativity, communication, critical thinking, and collaboration—that are deemed the most important skills for the 21st century by the Partnership for 21st Century Skills (2007), as well as the 6Cs proposed by Canadian education thinker Michael Fullan: character, citizenship, communication, critical thinking and problem-solving, collaboration and teamwork, and creativity and imagination (Fullan, 2013). Pennsylvania State psychologist Angela Duckworth champions the importance of grit (e.g., Duckworth et al., 2007), while Yale’s Zorana Ivcevic and Marc Brackett suggest personality traits matter more (Ivcevic & Brackett, 2014). Mind-set matters a lot, according to Stanford psychologist Carol Dweck (Dweck, 2008). Yong Zhao of the University Kansas suggests that the entrepreneurial mindedness as necessary (Zhao, 2012). And there are more: dispositions (Costa & Kallick, 2013), emotional intelligence (Goleman, 1995), self-determination (Ryan & Deci, 2000; Wehmeyer, Shogren, Little, & Lopez, 2017), and the seven survival skills for the future identified by Harvard’s Tony Wagner (Wagner, 2008, 2012).

No single proposal is the end-all solution to the wicked problem. Instead, what makes a person successful is ultimately a complex and unique combination of all the skills and competencies included in the above proposals (Rose, 2016; Zhao, 2018a). It is unlikely that any single ability or quality can determine the fate of a person’s success in the future (Goleman, 1995; Sternberg, 1996). It is likely that all or most of the proposed skills, knowledge, dispositions, motivations, or
personal traits play a role in more positive student outcomes. Thus, any limited selection of these would capture some and miss others at the same time.

**Wicked problems: Negative side effects**

Solutions to wicked problems always carry unforeseen or unforeseeable consequences because so many factors work at the same time and they are highly interactive. When a solution is applied, it may solve part of the problem, but it may also create new problems. Even worse, they could leave a host of negative side effects on students and the society (Zhao, 2017, 2018b).

For example, the almost exclusive focus on testing math and literacy from *No Child Left Behind* Act (*NCLB*) resulted in many unforeseen problems, including a narrowing of curriculum, cheating, and demoralization of educators (Ladd, 2017; Nichols & Berliner, 2007). While the *NCLB* law aimed to close the achievement gap, ultimately it shortchanged many children, narrowed the gap of schooling, and stifled educational innovation (Hess, 2011).

**Wicked problems: No immediate or ultimate test of a solution**

Another characteristic of wicked problems is that there is no immediate and ultimate test of the impact of the solution. In the case of deciding what to measure in education, we cannot immediately know if what is measured really matters, nor can we ultimately verify if what is measured actually matters in the future. Although we can use past evidence to infer what will matter in the future, there is no assurance that what matters now or in the past will matter equally in the future. Moreover, the concept of *success* varies a great deal, making it impossible to test fully the validity of any measured items in relation to future success.

The predictive power of the commonly used measures in education has been very limited. IQ scores, grade point average, and standardized tests such as the SAT and ACT have not been able to predict life success or even very narrow definition of success: college academic success (Zhao, 2016c). Early reading has been found to be associated with early education success (The Annie E. Casey Foundation, 2013), but it has also been found to be “associated with worse long-term outcomes including less overall educational attainment, worse teenage and adult adjustment, and increased alcohol use” (Kern & Friedman, 2009, p. 428). This is why there is increasing interests in finding more constructs to measure—to better capture the range of factors that influence success.

In essence, solutions to wicked problems are a bet, which cannot be proven right or wrong. Thus, when faced with a wicked problem, it is all the more important to ensure that the solution is as good as possible to begin with. Coming up with as-good-as-possible solutions for wicked problems requires unconventional strategies.
A collaborative approach to tackling wicked problems

There are different approaches for tackling wicked problems (Brown, Harris, & Russell, 2010; Roberts, 2000). According to Roberts (2000), there are three primary approaches to solving wicked problems. The authoritative approach is a top-down process through which decisions can be made quickly by a designated group of experts. By definition, the authoritative approach has misalignments with wicked problems, including not being able to identify recognizable and widely accepted experts. Thus, while this approach has the potential to bring an immediate, ideally workable solution, this solution causes disagreement and has difficulty gaining wide acceptance within the field of practice.

On the other hand, the competitive approach for tackling a wicked problem has various groups competing for the winning solution. As compared to the authoritative approach, the competitive approach has the potential to support more idea generation, yet it also has the potential to reinforce warring factions within the solution. This can lead to extraneous problems, supporting greater unrest, and consuming needed resources. In the end, the competitive approach may or may not produce workable and accepted solutions.

Finally, the collaborative approach to a wicked problem focuses on bringing various stakeholder teams and ideas together to work toward agreed-upon solutions. In the short term, the collaborative approach may seem slower and less ordered than the authoritative and competitive approaches. In the end, however, the collaborative approach has the potential to provide a host of workable solutions that are more widely accepted across the field. The key to supporting a collaborative approach is to establish purposeful support structures and agreed upon processes in place for encouraging open problem-solving.

Too often, deciding what matters in education has followed the authoritative approach. The decision about what to measure or teach has often been decided by a group of supposedly representative experts convened by government or education agencies or a local school board. Decisions about how to measure the what have followed the competitive approach—inviting different groups to bid for contracts. This dual approach is the case in the Common Core State Standards Initiative, where the standards were created by delegated individuals and the tests were developed by competing groups. It followed the traditional analytic, linear approach: defining the problem, coming up with a solution, inviting comments, and implementing. As an aside, the Common Core State Standards Initiative did not treat the decision of what matters in education as a wicked problem, and in the end, neither of these approaches has supported meaningful and lasting change in educational measurement or practice.

A collaborative approach is considered better in the field. This approach requires all who are impacted by the problem to actively participate in formulating solutions. Stakeholders of education
outcomes include students, parents, teachers, school leaders, employers, the public, and policy makers. In deciding what matters in education, all these stakeholders should be given the opportunity to contribute to developing workable solutions. Notably, students, parents, and teachers—the three groups of stakeholders with the most at stake in the solution—have traditionally played only marginal roles in what have been top-down approaches to school reform (Sarason, 1990) and must be supported to be involved as meaningful, active contributors to any solution to the *what matters* question.

**Focal questions**

In the collaborative process, stakeholders can have different opinions, but it would be more productive to focus the exchange of opinions on a set of meaningful questions. The tradition of deciding educational outcomes follows a winner-takes-all approach. That is, the prevailing opinion is applied to all students. In other words, whichever side wins the argument gets to decide and consequently impose the decided set as expected outcomes for all children. The winning set of outcomes become codified as curriculum standards, accountability measures for schools and teachers, and bases for high-stakes decisions about the life of students, for example, college admissions, grades retention, or designation for special or gifted education status.

As a result, the debates about what to count as outcomes that matter in education have been fierce, with different side working hard to convince the others and policy makers as well as the public that their proposed set has more merits than those of others. However, given the wickedness of the problem as discussed before, the disputes cannot be settled this way. Instead, we should challenge the winner-takes-all mind-set. Instead of asking which proposed set is superior, we can frame the debates in a more productive way with different questions:

- Is it necessary for all individuals to be equipped with all the proposed qualities to be successful, however defined?
- Is it possible for all children to acquire all the qualities?
- Are all qualities of equal importance all the time?
- Can all proposed qualities be measured? If not, are the unmeasured qualities less important?

**Do all children need the same set of qualities?**

Traditionally, once what matters is decided, it is applied to all children. There is often the assumption that all children need the same set of capabilities to succeed in life. But, upon close examination, this assumption may be mistaken. It is without question that everyone needs a set of basic *floor* qualities to perform the essential functions of being a member of common society. Although the set of floor qualities differs from society to society, and from time to time, it is very minimal
and should be a very narrow set of qualities. Typically, the floor qualities include basic literacy and numeracy and also some knowledge and skills to perform functions as a citizen. While the basic set may be the same as the qualities that enables one to live a successful life in some societies, more often than not, the minimal floor qualities cannot lead to a successful, financially independent, and psychologically fulfilling life in modern societies. Thus, a different set of qualities is needed beyond the basic set of minimal qualities.

The foundation of modern economy—the division of labor—demands specialization. As a result of the Fourth Industrial Revolution (Schwab, 2015; World Economic Forum, 2016) with technology performing even more complex but repetitive, identical tasks, the need for more specialization has grown even more acute. What makes an individual successful is often a unique combination of qualities that may not be replicated in another person or smart machine. The combination includes cognitive skills, noncognitive skills, and domain-specific knowledge and abilities. There is no one profile of qualities that is universally applicable to all tasks, jobs, and professions (Rose, 2016; Zhao, 2012, 2016b, 2018a), nor is it technically possible for all students to develop the same qualities (Barrett, 2017). For example, what makes a musician successful is certainly quite different from the qualities that help an engineer perform well. Even within a profession, there are different tasks, for example, there are routine engineering jobs that require different qualities from creative engineering jobs. Research has found that creative individuals in fine and performing arts have different profiles of personalities and skills from creative individuals in technical/engineering fields (Kerr & McKay, 2013).

Basic primary and secondary education have been tasked with equipping children with both sets of qualities—floor and ceiling qualities. But there is considerable confusion over what are floor qualities and ceiling qualities. Ceiling qualities that can make individuals successful are often viewed the same as the basic qualities that everyone should develop. As a result, schools are asked to require all individuals to develop the same set of qualities. For example, international assessment programs such as the Programme for International Student Assessment (PISA), a test of 15-year-olds’ competencies in reading, math, and science, suggest that all students in the world need the same set to succeed in the 21st century. Similarly, the U.S. CCR prescribes the same set of qualities deemed necessary for success after secondary school for all American students.

Hence, discussions about what matters in education need to focus on what is the minimal set of qualities that everyone should acquire and how much effort should be devoted to ensuring that all children acquire the same set of basic qualities. Once answered, then discussions can focus on other qualities. For example, when should children be given the opportunities to specialize to develop their unique set of qualities? What kind of exceptions should be given to students in special circumstances when demanding the basic qualities may hinder their long-term development?
Is it possible to pursue all proposed outcomes?

Another important question to consider when attempting to resolve the wicked problem of what matters in education is whether it is desirable to pursue all proposed outcomes, even if it were possible for all children to be equipped with all the qualities. The answer to this question lies with an understanding of the relationships among the different qualities.

Ecology provides a useful framework for understanding the relationships among the different outcomes. Ecologists have identified five important types of interactions between two organisms: (a) competition—both organisms have some kind of negative effect on each other; (b) predation—positive for one (the predator) and negative for the other (the prey); (c) parasitism—negative for one (the host) and positive for the other (the parasite); (d) commensalisms—positive for one (the commensal) and no effect on the other; and (e) mutualism—positive for both (Odum, 1997). We can imagine individual students as an ecosystem in which the different qualities interact with each other as organisms. The different qualities (what matters) can be imagined to have similar types of relationships as living organisms in an ecosystem. There are perhaps four types of relationships that exist among the different education outcomes or intended qualities.

**Competition**

Two qualities compete with each other for resources. In an ecosystem, lizards and frogs, for example, are in competition because they both eat small insects. This is a win–lose relationship. This relationship exists among educational outcomes all the time. For example, different subjects are in constant competition for time and other instructional resources, as well as students’ attention. A student cannot simultaneously be devoting time to music and math at the same time because time is constant. For the same reason, a school cannot possibly increase time for math or reading without taking time away from other activities. Increasing time for one subject necessarily reduces time for other subjects. This win–lose relationship was evidenced in the effect of NCLB—schools increased time for math and reading, which resulted in taking time away from other subjects and activities, such as social studies, science, arts, music, and even recess (King & Zucker, 2005; McMurrer, 2007).

**Predation**

Desired qualities can also have a predatory relationship. In ecosystems, a predatory relationship is one in which the growth of the predator relies on the disappearance of the prey. This is also a win–lose relationship. For instance, birds gain energy by eating earthworms, but earthworms gain no benefit. Predatory relationships exist among educational outcomes because the growth of some outcomes depends on the decrease of others. For example, increased levels of obedience and compliance rely on reduction in the willingness to question the status quo and authority or to
express one’s own thoughts and opinions. An individual cannot be compliant and creative at the same time. Academic performance typically reflects one’s willingness to follow instructions and provide predetermined answers, while creativity reflects more on one’s confidence and courage to question the status quo and express one’s own views. There is evidence (e.g., Pretz & Kaufman, 2017) suggesting that high levels of academic performance in the form of high school class rank can come at the cost of creative confidence. Thus, educational strategies that focus on increasing academic performance can prey on creative expression.

**Commensalism**

Some qualities can benefit from other outcomes without benefiting or harming these other outcomes, a commensalism relationship. The transparent shrimp, for example, lives in a reef that provides benefits to the shrimp, such as camouflage, but the reef does not receive any benefit or damage from this relationship. This is a win–neutral relationship. In education, such relationships exist as well because the improvement in some outcomes is dependent on the increase of others, but the relationship is unidirectional. For example, evidence suggests that grit and growth mind-set can improve academic outcomes (Claro, Paunesku, & Dweck, 2016; Duckworth et al., 2007), but there is little evidence that academic outcomes also increase or decrease grit or growth mind-set.

**Mutualism**

It is also possible for some qualities to have a relationship in which they benefit from interacting with one another. An example of this type of relationship in ecology is that of bees and flowers in which bees get nectar from flowers and in return spread pollen so the plant can reproduce. This is a win–win relationship. In education, mutually beneficial relationships exist among outcomes as well. For instance, it is possible that self-determination and emotional well-being are mutually enhancing. When one is able to experience more autonomy, he or she has increased sense of well-being psychologically (Ryan & Deci, 2017).

The types of relationships between qualities are presented in Table 1. Of the four types of possible relationships, one is mutually enhancing, which means improving one quality can

| Quality X | Quality Y |
|-----------|-----------|
| Competition | +         | –          |
| Predation   | +         | –          |
| Commensalism | +         | 0          |
| Mutualism   | +         | +          |

**Table 1. Types of relationships between two qualities.**
improve another. One is commensalism, meaning that efforts to increase one quality do not help or hurt the others. Two types of relationships indicate that an intervention or change intended to improve one quality can hurt the development of another, resulting in negative relationship.

The negative relationship is what we need to pay attention to, and there is evidence that confirms the negative relationship between some outcomes. For example, international assessment programs such as the PISA and Trends in International Mathematics and Science Study (TIMSS) show a negative correlation between test scores and students’ confidence in enjoyment and value of learning (Loveless, 2006; Zhao, 2017). Talent and grit may have a negative relationship as well (Perkins-Gough, 2013). Studies have also shown that short-term instructional outcomes may come at the cost of long-term outcomes. For example, explicitly teaching students may result in immediate success in learning target content, but it can cause a loss of curiosity, creativity, and deep understanding (Bonawitz et al., 2011; Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Kapur, 2014, 2016; Kapur & Bielaczyc, 2012; Peterson, 1979).

The existence of negative relationships between qualities suggests that an individual cannot possibly pursue all qualities equally successful. It is thus unreasonable to hold schools and teachers accountable for ensuring that all students be equipped with all the proposed qualities. Is it necessary, then, for stakeholders to consider what qualities are important? Do we want creative and curious children or high scores on standardized tests? Do we want children to be good at math and literacy at the cost of music and arts?

Moreover, stakeholders may also reimagine an education system that would support individuals to pursue their own interest and strength as a way to develop their unique profiles of qualities. In other words, if it is unnecessary and impossible for individuals to acquire the same set of qualities, schools can create an environment that supports each individual to develop their strengths and passions (Zhao, 2018a). As a result, measuring what matters becomes measuring what matters for individuals instead of the average of groups of individuals (Basham, Hall, Carter, & Stahl, 2016).

Are outcomes outcomes?

There is much confusion about outcomes. In the long list of proposed outcomes, some are treated as input variables sometimes; at other times, they are treated as output variables. For example, growth mind-set is often treated as an input variable that can enhance academic achievement, so are grit, confidence, and mental health. But should these be educational outcomes in their own right? At the same time, academic outcomes, such as grades, test scores, and education attainment, have been treated as output variables, but they can negatively or positively affect mind-set, confidence, grit, and mental health. A more productive approach is to treat all outcomes as both input and output variables at the same time—then causality is bidirectional.
This approach is especially important considering that some of the outcomes are more short term than others. For example, academic achievement is often a measure of short-term effects of schooling. Although there are measures of long-term cumulative effect of schooling in some academic domains, such as the ACT, PISA, and TIMSS, most of the time academic achievement is a measure of the degree to which students have mastered the intended knowledge and skills within a relatively short period of time. Grades, one of the most commonly used measures of academic achievement, are typically given at the end of a course. Moreover, grades are based on even smaller units of measurement, such as weekly quizzes, end of an instructional unit exam, daily homework completion, and midterm and end-of-course exams. Standardized tests are often given annually in a limited number of subjects.

In research, longitudinal studies of academic achievement are not common and short-term academic achievement has had more influence on practice and policy. The effectiveness of interventions is more often than not judged by its effect on short-term academic achievement. Schools and teachers are held accountable for improving academic achievement in the short term, as exemplified in the annual academic measures discussed in Every Student Succeeds Act. Important decisions are made about students-based short-term academic performance. For example, students are prescribed remediation, grade retention, special education, or gifted education based on performance over a year or scores on a test. Furthermore, parents judge their children and their children’s education based on improvement of academic achievement over a short period of time. Consequently, short-term academic achievement drives actions in education.

Efforts to boost short-term academic outcome can negatively affect important long-term outcomes. There is emerging evidence that short-term positive academic outcomes do not necessarily translate into long-term life quality outcomes. For example, the longitudinal study by Howard Friedman found that early literacy was negatively associated with important indicators of life quality such as social emotional well-being and adjustment in long term (Kern & Friedman, 2009).

There is also evidence that instructional interventions that resulted in more effective information acquisition or imitation can cause a suppression of curiosity and creativity (Bonawitza et al., 2011; Buchsbauma et al., 2011). Researchers have also found that extracurricular activities tended to be a stronger predictor of creative expression in college applicants than traditional admissions factors, such as SAT scores and high school rank (Cotter, Pretz, & Kaufman, 2016).

A focus on short-term academic outcomes, then, can negatively affect long-term academic outcomes. For example, research has found that teaching decoding skills does not improve reading comprehension; however, decoding skills can be viewed by some as necessary part of reading achievement. Thus, efforts and time devoted to learning decoding skills is time and efforts away from actually learning to read (Zhao, 2018b). Some strategies that teach children to memorize facts can show immediate positive effect, but they may negatively damage children’s interest in the
subject or constrain them from developing deeper conceptual understanding of the subjects (Kapur, 2014, 2016).

It is thus important for stakeholders to consider the short-term outcomes in light of possible long-term outcomes. Is it worth it, for example, to make sure students memorize the multiplication table at the cost of their losing interest in math? Or is it worth it if they passed tests in a course but developed a negative attitude to the subject? Furthermore, is any knowledge acquisition as important as maintaining a creative inquisitive mind-set?

Conclusions
In summary, we believe that deciding what matters is a wicked problem and should be treated as such. Wicked problems are drastically different from tame or technical problems and thus require unconventional approaches. Historically, education has used both authoritative and competitive approaches to solve problems. Unfortunately, these approaches have supported little to no meaningful lasting impact on the education system. It is time to reflect on what matters in education, then in turn what should be measured, and eventually how to measure it. System-level reflection on how to support massive collaboration in other areas has the potential to provide a structure of this type of collaboration in education. Baselining this collaboration with a basic set of underlying guiding questions can start this process. We hope a collaborative approach suggested in this article can result in a better solution to this wicked problem.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References
Aspen Youth Entrepreneurship Strategy Group. (2008). Youth entrepreneurship education in America: A policy maker’s action guide. Retrieved from http://www.entre-week.org/eweek_files/YouthEntrepreneurshipinAmericaYESG_report[4].pdf
Australian Public Service Commission. (2012). Tackling wicked problems: A public policy perspective. Retrieved from http://www.apsc.gov.au/publications-and-media/archive/publications-archive/tackling-wicked-problems
Barrett, M. (2017). Early lexical development. In P. Fletcher & B. MacWhinney (Eds.), The Handbook of child language (361–392). Hoboken, NJ: Blackwell.
Basham, J. D., Hall, T. E., Carter Jr., R. A., & Stahl, W. M. (2016). An operationalized understanding of personalized learning. Journal of Special Education Technology, 31, 126–136.
Beghetto, R. A. (2017). Legacy projects: Helping young people respond productively to the challenges of a changing world. Roeper Review, 39, 1–4.

Bonawitz, E., Shaftob, P., Gweon, H., Goodmand, N. D., Spelkee, E., & Schulze, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. Cognition, 120, 322–330.

Brown, V. A., Harris, J. A., & Russell, J. Y. (2010). Tackling wicked problems: Through the transdisciplinary imagination. New York, NY: Earthscan.

Brunello, G., & Schlotter, M. (2010). The effect of non-cognitive skills and personality traits on labour market outcomes. Retrieved from http://www.epis.pt/downloads/dest_15_10_2010.pdf

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shaftob, P. (2011). Children’s imitation of causal action sequences is influenced by statistical and pedagogical evidence. Cognition, 120, 331–340.

Cheng, L., & Watanabe, Y. (2004). Washback in language testing: Research contexts and methods. New York, NY: Routledge.

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. Proceedings of the National Academy of Sciences, 113, 8664–8668.

Conklin, J. (2006). Dialogue mapping: Building shared understanding of wicked problems. Chichester, England: Wiley.

Costa, A. L. L., & Kallick, B. (2013). Dispositions: Reframing teaching and learning. Thousand Oaks, CA: Corwin.

Cotter, K. N., Pretz, J. E., & Kaufman, J. C. (2016). Applicant extracurricular involvement predicts creativity better than traditional admissions factors. Psychology of Aesthetics, Creativity, and the Arts, 10, 2.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92, 1087–1101.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. Educational Researcher, 44, 237–251.

Dweck, C. S. (2006). Mindset: The new psychology of success (1st ed.). New York, NY: Random House.

Dweck, C. S. (2008). Mindset: The new psychology of success (Ballantine Books trade paper back ed.). New York, NY: Ballantine Books.

European Communities. (2006). Key competences for lifelong learning: A European framework. Retrieved from Luxemburg https://www.voced.edu.au/content/ngv%3A59967

Fullan, M. (2013). Great to excellent: Launching the next stage of Ontario’s education agenda. Retrieved from http://www.michaelfullan.ca/wp-content/uploads/2013/09/13_Fullan_Great-to-Excellent.pdf

Goleman, D. (1995). Emotional intelligence. New York, NY: Bantam Books.

Hess, F. M. (2011). Our achievement-gap mania. National Affairs, Fall, 2011, 113–129.

Hirsch, E. D., Jr. (2010). How to save the schools. The New York Review of Books. Retrieved from https://www.nybooks.com/articles/2010/05/13/how-save-schools/

Ivcevic, Z., & Brackett, M. (2014). Predicting school success: Comparing conscientiousness, grit, and emotion regulation ability. Journal of Research in Personality, 52, 29–36.

Kapur, M. (2014). Productive failure in learning math. Cognitive Science, 38, 1008–1022.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51, 289–299.
Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*, 45–83.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology, 13*, 1–12.

Kern, M. L., & Friedman, H. S. (2009). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology, 30*, 419–430.

Kerr, B., & McKay, R. (2013). Searching for tomorrow’s innovators: Profiling creative adolescents. *Creativity Research Journal, 25*, 21–32.

King, K. V., & Zucker, S. (2005). Curriculum narrowing. *Policy Report*. Retrieved from http://www.pearsonassessments.com/NR/rdo/dores/D3362EDE-7F34-447E-ADE4-D4CB2518C2B2/0/curriculumnarrowing.pdf

Ladd, H. F. (2017). No child left behind: A deeply flawed federal policy. *Journal of Policy Analysis and Management, 36*, 461–469.

Levin, H. M. (2012). More than just test scores. *Prospects: The Quarterly Review of Comparative Education, 42*, 269–284.

Loveless, T. (2006). *How well are American students learning?* Retrieved from http://www.brookings.edu~/media/Files/rc/reports/2006/10education_loveless/10education_loveless.pdf

McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Retrieved from http://www.cepp-dc.org/displayDocument.cfm?DocumentID=312

Mishkind, A. (2014). *Overview: State definitions of College and Career Readiness*. Retrieved from https://ccrscenter.org/sites/default/files/CCRS%20Definitions%20Brief_REV_1.pdf

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America’s schools*. Cambridge, MA: Harvard Education Press.

Norris, N. D. (2004). *The promise and failure of progressive education*. London, England: Scarecrow Education.

Odum, E. P. (1997). *Ecology: A bridge between science and society*. Sunderland, MA: Sinauer Associates.

Partnership for 21st Century Skills. (2007). *Framework for 21st century learning*. Retrieved from http://www.21stcenturyskills.org/documents/frameworkflyer_072307.pdf

Perkins-Gough, D. (2013). The significance of grit: A conversation with Angela Lee Duckworth. *Educational Leadership, 71*, 14–20.

Peterson, P. L. (1979). Direct instruction: Effective for what and for whom. *Educational Leadership, 37*, 46–48.

Pretz, J. E., & Kaufman, J. C. (2017). Do traditional admissions criteria reflect applicant creativity? *The Journal of Creative Behavior, 51*, 240–251.

Ravitch, D. (2001). *Left back: A century of battles over school reform*. New York, NY: Simon & Schuster.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.

Roberts, N. (2000). Wicked problems and network approaches to resolution. *International Public Management Review, 1*, 1–19.

Rose, T. (2016). *The end of average: How we succeed in a world that values sameness* (1st ed.). New York, NY: HarperOne.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York, NY: Guilford.

Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it’s too late?* The Jossey-Bass Education Series and the Jossey-Bass Social and Behavioral Science Series. San Francisco, CA: Jossey-Bass.

Schwab, K. (2015, December 12). The fourth industrial revolution: What it means and how to respond. *Foreign Affairs*.

Seligman, M. E. P. (2006). *Learned optimism: How to change your mind and your life*. New York, NY: Random House.

Seligman, M. E. P. (2011). *Flourish: A visionary new understanding of happiness and well-being* (1st Free Press hardcover ed.). New York, NY: Free Press.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*, 298–317.

Sternberg, R. J. (1996). *Successful intelligence: How practical and creative intelligence determine success in life*. New York, NY: Simon & Schuster.

The Annie E. Casey Foundation. (2013). *Early warning confirmed: A research update on third grade reading*. Retrieved from http://www.aecf.org/m/resourcedoc/AECF-EarlyWarningConfirmed-2013.pdf

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: John Wiley.

Wagner, T. (2008). *The global achievement gap: Why even our best schools don’t teach the new survival skills our children need—And what we can do about it*. New York, NY: Basic Books.

Wagner, T. (2012). *Creating innovators: The making of young people who will change the world*. New York, NY: Scribner.

Wehmeyer, M. L., Shogren, K. A., Little, T. D., & Lopez, S. J. (Eds.). (2017). *Development of self-determination through the life-course*. Berlin, Germany: Springer.

Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research, 61*, 1–24.

World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the Fourth Industrial Revolution*. Retrieved from http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Wraga, W. G. (2001). Left out: The villainization of progressive education in the United States. *Educational Researcher, 30*, 34–39.

Zhao, Y. (2012). *World class learners: Educating creative and entrepreneurial students*. Thousand Oaks, CA: Corwin.

Zhao, Y. (2016a). *Counting what counts: Reframing education outcomes*. Bloomington, IN: Solution Tree Press.

Zhao, Y. (2016b). From deficiency to strength: Shifting the mindset about education inequality. *Journal of Social Issues, 72*, 716–735.

Zhao, Y. (2016c). Numbers can lie: The meaning and limitations of test scores. In Y. Zhao (Ed.), *Counting what counts: Reframing education outcomes* (pp. 13–30). Bloomington, IN: Solution Tree Press.
Zhao, Y. (2017). What works can hurt: Side effects in education. *Journal of Educational Change, 18*, 1–19.
Zhao, Y. (2018a). *Reach for greatness: Personalizable education for all children*. Thousand Oaks, CA: Corwin.
Zhao, Y. (2018b). *What works can hurt: Side effects in education*. New York, NY: Teachers College Press.