Narrow Assessments Misrepresent Development and Misguide Policy

Comment on Steinberg, Cauffman, Woolard, Graham, and Banich (2009)

Kurt W. Fischer, Zachary Stein, and Katie Heikkinen

Harvard Graduate School of Education

Intellectual and psychosocial functioning develop along complex learning pathways. Steinberg, Cauffman, Woolard, Graham, and Banich (2009) measured these two classes of abilities with narrow, biased assessments that captured only a segment of each pathway and created misleading age patterns based on ceiling and floor effects. It is a simple matter to shift the assessments to produce the opposite pattern, with cognitive abilities appearing to develop well into adulthood and psychosocial abilities appearing to stop developing at age 16. Their measures also lacked a realistic connection to the lived behaviors of adolescents, abstracting too far from messy realities and thus lacking ecological validity and the nuanced portrait that the authors called for. A drastically different approach to assessing development is required that (a) includes the full age-related range of relevant abilities instead of a truncated set and (b) examines the variability and contextual dependence of abilities relevant to the topics of murder and abortion.

Keywords: adolescence, assessment, cognitive development, learning pathway, maturity

Children develop many different capabilities along multiple pathways, and they gradually learn to use those capabilities across many situations and in many different emotional states. Steinberg, Cauffman, Woolard, Graham, and Banich (2009, this issue) oversimplified development by dividing it into two categories—cognitive and psychosocial—and measuring behavior in a few tasks that did not capture the full range of capabilities and contexts relevant to the issues of abortion and murder. Within the full range, children develop many cognitive and psychosocial capabilities long before age 16, and they continue to develop others in both areas long after age 16. The simplifications in the research by Steinberg and colleagues omit consideration of the breadth of cognitive and psychosocial capabilities pertinent to decisions about abortion and murder and thus lead to conclusions about policy that are misleading.

Consider Sally, a 17-year-old adolescent who has lived with her mother since she was 7, when her father moved out after a divorce. Her behavior illustrates the range of capabilities relevant to the actions of children and adolescents as well as the importance of contexts for shaping those actions. At age 17, Sally’s relationship with her mother has become strained. She struggles to control herself around her mother. Outbursts and arguments are almost daily occurrences. Sally knows that these irrational arguments typically end in tears and frustration, but she cannot seem to control herself. She wonders why she cannot control her emotions at home the way that she did when she was younger, when during her parents’ divorce she could “hold herself together” and handle difficult situations when her mother was in distress.

Sally’s situation at school is different from the one at home. During her parents’ divorce, when she was 6 and 7 years old, she struggled in school, and teachers suggested that she be held back a year so her skills could develop to the appropriate level. But now in high school, she excels in many classes and serves as a peer mediator, volunteering her time to help her classmates resolve their conflicts. Sally wonders how she can be so together at school now when she feels so out of control at home. She finds it curious that years ago she was a mess in school but was more capable of holding herself together at home.

Sally’s story exemplifies the variability and complexity of cognitive, emotional, and interpersonal development during childhood and adolescence. In some respects, Sally can be viewed as more emotionally mature and psychosocially capable at the age of 7—when the situation demanded it—than she is now at the age of 17. On the other hand, her cognitive capabilities seem to display the opposite trend, changing from below average in grade school to excellent in high school. In other words, her interpersonal capabilities—as well as her cognitive and emotional ones—vary or fluctuate drastically depending on context, as she indulges in emotionally charged, irrational arguments at home while serving as a peer mediator at school.
Embracing the Complexity of Development

This is how development looks during childhood and adolescence. Capabilities develop with age and experience, and they also fluctuate over time and across contexts and domains. Most broad capabilities—cognitive, emotional, and interpersonal—begin to develop during infancy and are dynamically constructed well into adulthood. Development of new capabilities continues far beyond the age of 16. At all ages, development of new capabilities is marked by discontinuities, spurts, and regressions and can unfold along diverse pathways as various lines of development interact. For over a century, psychologists and theorists have attempted to embrace and explain the complexity and diversity of human development (Baldwin, 1894, 1906; Fischer & Bidell, 2006; Piaget, in Gruber & Voneche, 1977; Wilber, 1999).

When the full complexity of human development is considered, it is hard to make simple arguments and offer clear-cut advice. Steinberg and colleagues (2009) chose to jettison this complexity—both conceptually and methodologically—in order to make a simple argument and offer clear-cut advice for policy. They measured capabilities by using tasks that are known to have ceiling effects at specific ages and that have limited ecological validity (little relation to the everyday lives of adolescents). Thus, they did not capture the broad spectrum of related and lived abilities that constitute cognitive, emotional, and interpersonal development—assessing instead only part of the pathways for “intellectual” and “psychosocial” development. For intellectual capabilities, they assessed skills that develop early, while for psychosocial capabilities, they focused on skills that develop later, although they did not intend to tie their measures to specific ages a priori. They conceptualized the development of capabilities in essentialist and nondynamic terms, ignored the radical importance of context, and assumed a static picture in which a capability, once it has been displayed, is taken as simply present.

We examine their perspective on the development of capabilities in adolescence in terms of two basic research issues: construct validity (Did they conceive and measure their basic constructs in a psychologically meaningful way?) and ecological validity (Were they measuring constructs that are meaningfully related to the real lives of adolescents, especially the complex issues related to abortion and murder?). It is important to note that people develop differently in separate domains on the basis of their experiences and interests, and their development continues far into adulthood in domains on which they focus (Dawson-Tunik, 2004; Fischer, Yan, & Stewart, 2003; Kitchener, Lynch, Fischer, & Wood, 1993; Kohlberg, 1984).

Measuring Development in Childhood and Adolescence: Construct Validity

Steinberg and colleagues (2009) sought to justify the “flip-flop” in which the American Psychological Association maintained that teens are mature enough to make reasonable decisions about abortion but not mature enough to be treated as adults in death penalty cases. They argued that this apparent flip-flop is justified by research findings that teens attain adult-like capabilities at different ages in different domains. Specifically, they argued that adolescents’ cognitive capabilities develop earlier than their capacity for emotional self-regulation and impulse control. To support this claim, they reported research using a battery of tasks to measure the proposed differential distribution of capabilities.

Cognitive Capabilities

To assess cognitive maturity, Steinberg et al. (2009) administered “tests of basic intellectual functioning” (p. 587), which included tests of working memory interference, digit span memory, and verbal fluency. We discuss these tests’ construct validity and their relevance to the larger construct of “cognitive capacity,” or “cognitive maturity.” Although the authors acknowledged that theirs was an “incomplete measure of cognitive capacity” (p. 590), they nonetheless made the bold claim that general cognitive capacity plateaus around age 16. This claim is spurious on at least two grounds: (a) These tasks do not adequately represent general cognitive capacity, and (b) cognitive capacity, when more carefully defined, does not plateau around age 16 or 17. The authors used tasks with ceiling effects.

Attempts to measure something like general cognitive capacity with a single metric or composite go back to Binet’s laboratory in Paris at the turn of the 20th century, where intelligence testing was invented. Although Steinberg and colleagues (2009) separately assessed IQ in order to control for it, their measures of “basic intellectual functioning” are quite close to traditional measures of intelligence. Even the best tests of intelligence do not capture
everything about intelligence (Gardner, 1983), let alone a broader construct like cognitive capacity. Working-memory interference tasks, digit-span memory tasks, and tests of verbal fluency—the tasks used by Steinberg et al.—are known to peak or plateau in the teenage years, and they constitute only a few of the skills involved in cognitive development (Horn, 1982).

Cognitive development includes many interrelated and complex skills, such as logical reasoning (Inhelder & Piaget, 1958), reflective judgment (Kitchener et al., 1993), conceptual complexity (Case, 1992; Fischer & Bidell, 2006), and emotion regulation (Benes, Turtle, Khan, & Farol, 1994; Dawson-Tunik, 2004; Fischer et al., 2003; Vaillant, 1977). These skills all begin to develop early in life, continue to develop well into adulthood, and vary as a function of learning, context, and maturation. In contrast, the standard tests of working memory and verbal fluency are simpler tasks that plateau early. More complex skills such as reflective judgment, logical reasoning, and even working memory for sophisticated concepts such as conservation of energy, evolution by natural selection, and the role of community in development do not plateau in the teenage years.

The evidence is substantial that cognitive development continues for years beyond the ages of 15 or 16 (Dawson-Tunik, 2004; Fischer et al., 2003), and likewise brain development continues well into adulthood (Benes et al., 1994). Steinberg and colleagues (2009) were not justified in claiming that “by age 16, adolescents’ general cognitive abilities are essentially indistinguishable from those of adults” (p. 592). Their research indicators of working memory and verbal fluency may have reached adult levels by that age, but complex and important cognitive capabilities continue to develop long afterward. Even vocabulary continues to increase throughout adulthood, especially for educated people (Horn, 1982). The measures used by Steinberg et al. fail to capture the broader developmental patterns that characterize adolescent and adult cognition.

An accurate portrait of development requires assessment of a diverse set of skills that begin early in life and move along learning pathways into adulthood (Case, 1992; Fischer & Bidell, 2006). Some of these skills, such as standard assessments of object permanence and working memory, develop quickly and plateau early, while others, such as reflective judgment, start early but do not plateau until well into adulthood (Kitchener et al., 1993). Skill-specific learning sequences develop at different rates depending on a variety of factors, including education (Dawson-Tunik, 2004), contextual support (Fischer & Bidell, 2006), and motivation (Fink, 2007). The development of cognitive maturity thus varies depending on what is measured and how it is measured. Simpler skills such as working memory for common words or numbers and understanding of the basic requirements of participation in a trial (testifying honestly, sitting quietly, etc.) may peak by adolescence, but complex conceptual skills such as reflective judgment, understanding how laws and courts function in society, and taking perspectives across different cultures do not peak until the 20s or 30s or later—and even then only in the context of optimal contextual support and education.

Psychosocial Capabilities

The same problems hold for Steinberg et al.’s (2009) approach to psychosocial capabilities. The measures in their research represent only a small segment of social and emotional skills and thus distort the picture for development of these capabilities. To assess “psychosocial maturity” (p. 588), Steinberg et al. collected self-report questionnaires on risk perception, sensation seeking, impulsivity, peer influence, and future orientation. All measures assessed self-descriptions—not self-regulation, understanding about legal processes, moral judgment, or other capabilities relevant to abortion and murder.

With these instruments, they observed essentially no change in psychosocial scores, on average, between the ages of 10 and 17 years and then significant growth to ages 18–21 and age 26 and beyond. The problem, again, is that the measurement tasks truncate the construct and thus drastically constrain the age range in which development is observable. Different measures of psychosocial maturity display significant and important developmental changes at early ages, such as basic sociomoral perspective taking (Kohlberg, 1984; Selman & Schultz, 1990) and basic interpersonal attunement and attachment (Cassidy & Shaver, 1999). Children construct their understanding of self and others starting at birth and show richly textured developmental pathways for psychosocial skills from infancy through adulthood (Fischer & Bidell, 2006). Self-reports about risk, impulsivity, peer influence, and so forth represent a limited set of behaviors and do not capture fundamental capabilities.
depending on what is measured, the developmental pattern for psychosocial maturity will differ greatly. Self-report measures like those used in the Steinberg et al. (2009) study will show a late onset for psychosocial maturity. On the other hand, standard measures for development of attachment relationships or social perspective taking will show major transformations of psychosocial maturity at early ages. Infants in the first two years develop skills for self-regulation with their mothers or other caregivers. Grade-school children come to grasp the need for abstract norms and rules and the reasons for controlling their behavior to follow those norms and rules. These skills lay the foundation for understanding about abortion and murder, although sophisticated understanding certainly awaits development of higher capabilities.

The claim by Steinberg et al. (2009) that “psychosocial maturity” sets in later than “cognitive maturity” is dependent on the selective use of specific measures. Using a different set of measures could yield the opposite picture. For example, the development of abstract reasoning capabilities such as reflective judgment, understanding the legal system, and taking diverse cultural perspectives lags behind the development of rich emotional and interpersonal attunements and rule following evident in early childhood. This comparison would lead to the opposite conclusion from that reached by Steinberg et al.: Psychosocial capabilities develop richly in childhood and early adolescence (and continue to develop well into adulthood).

Dynamic Webs of Development

Development in any domain—for example, cognitive or psychosocial—moves at a varying pace along multiple strands in a dynamic web across the life course (Fischer et al., 2003). In addition, a capability is not fixed across contexts, but varies dynamically. Perspective taking with your mother is different from perspective taking with your friend, your classmate, or your lawyer. The idea is simply wrong that once a capability has been displayed at a certain level in a certain context it will be displayed at that same level across a variety of contexts. Development unfolds along diverse pathways and is radically sensitive to variations in context. Skills mastered in one context can fall apart in another, needing to be rebuilt to meet the unique task demands of new situations. Factors that drive such variation include stress (Ayoub & Fischer, 2006), novelty (Granott, 2002), and recalibration/self-organization (Van Geert, 1994).

This kind of sensitivity to context is particularly significant in the two examples given by Steinberg and colleagues (2009) because the contexts are so vastly different, as they noted. Reasoning about abortion, where a doctor or health-care worker can support the teen’s thinking over a length of time, is very different from acting violently in the heat of the moment. Teenagers’ capabilities are tied to contexts and emotional states. Teenagers are not simply cognitively mature and psychosocially immature. Context is radically implicated in the nature of capabilities, and the two cannot be realistically disentangled. Depending on context and support, the same individual can function in drastically different ways, and there is not one condition that represents the true capability.

How does this argument apply to Sally’s story? Is she cognitively mature or immature? Is she emotionally and socially mature or immature? Is she more mature cognitively or psychosocially? These kinds of questions are predicated on a simple and essentialist notion of what capabilities are. Sally is both immature and mature depending on both the context she is in and the measuring instruments used. At school, Sally scores high on some measures of cognitive capability, as reflected in her strong performance in her courses; but she is still years away from sophisticated reasoning about reflective judgment (the bases of knowledge about complicated issues). At home during the heat of an argument with her mother, Sally does not take multiple perspectives, as she does during school, but focuses primarily on her own immediate feelings. But when she is serving as a peer mediator, she effectively takes her peers’ perspective. What is required is a rich portrait of her capabilities in different contexts and for different goals.

Sally’s story is typical, not unusual. Variability is the norm at all ages and especially in adolescence. The way that development and capability are measured can distort as much as it can reveal. Steinberg and colleagues (2009) systematically misrepresented the development of the broad capabilities they studied because of their selection of measures and their framing of capabilities as fixed and stable, as opposed to variable and context dependent. With these oversimplifications, they have created a simple story for policymakers, but the story is wrong.
Measuring Things That Are Worth Measuring: Ecological Validity

Because of the many strands of the web of development, scientists need to be careful about the measures they use and the claims they make about the implications of their findings. One way of increasing the likelihood that research will connect to practical questions and policies is to ask about the ecological validity of measures and methods. How do they relate to the everyday lives and behaviors of adolescents, especially in contexts relevant to the practical questions?

The real-world behaviors of children and adolescents such as Sally are not merely the expression of certain general capabilities functioning at certain age-specific levels. They arise from the complex and unique lives of individual people, rich with emotion, diverse relationships, and novel challenges. What is the relation between the assessment tasks that Steinberg et al. (2009) used and the dilemmas of adolescent life they are meant to illuminate? How can performances of working memory interference, digit span memory, and verbal fluency indicate the capacity to make a reasoned decision about abortion? Likewise, how can self-report of risk perception, sensation seeking, impulsivity, resistance to peer influence, and future orientation indicate the capacity to make decisions about criminal behavior? Measures should assess the capabilities that are involved in those kinds of decisions.

Unfortunately, the kinds of tasks and research paradigms used in the Steinberg et al. (2009) research are common in psychology. They truncate and oversimplify the range and variability of behaviors while lending themselves to neat statistical analyses and simple stories that distort understanding of real children in living contexts. They lead to the mistaken belief that capabilities develop in lockstep, age-fixed sequences and show no variability from child to child or from context to context. They are divorced from the lived experiences, behaviors, and challenges that real people face.

The alternative to these forms of psychological research and measurement (and the essentialist constructs they engender) is the use of dynamic methods for studying the development of behaviors in medias res (Fischer & Bidell, 2006; Mascolo & Fischer, in press). (For more on these methods and the measures that facilitate them, see also http://www.lectica.info/; Dawson & Stein, 2008; Scher & Dawson, 2002.) In real life, people cope with the many complexities of life “in the middle of things.” They display their behaviors and reasoning capabilities across a variety of situations and conditions, dealing with difficult decisions and situations such as those involved in abortion and murder. Dynamic structural methods can detect pathways for learning and action in these situations and in related domains such as making sociomoral decisions and analyzing oneself and others in important relationships. These methods begin with a common scale (a general ruler) that captures the ways that people build skills and concepts in any domain. They analyze activities that relate to the key domains of interest, such as dilemmas, stories, and situations that deal with abortion or murder. They depict learning pathways and limits on capabilities that are based on people’s activities in the domains of interest. They do not assume that people have general capabilities that somehow apply effortlessly across situations and contexts. This kind of research can uncover the variety of adolescent developmental pathways and the patterns of variability across different contexts.

Conclusion: Narrow Assessments Create Results That Flip-Flop

The ways in which Steinberg and colleagues (2009) measured cognitive capacities and psychosocial maturity led directly to their results and to misleading conclusions. A simple shift in the assessments used would reverse the pattern, so that cognitive capacities would seem to develop during adulthood while psychosocial abilities would seem to develop during childhood, ending at age 15 or 16. Simply put, the measures in the study made the results come out in favor of the hypothesis. Moreover, the measures lacked a realistic connection to the lived behaviors of adolescents and thus lacked ecological validity. From a broad perspective of the full range of capabilities, both cognitive and psychosocial, development involves many skills that develop along complex pathways from infancy through adulthood, with many capabilities developing both before age 16 and afterward during early adulthood. These capabilities can be measured by observations of activities in relevant contexts and placed within developmental pathways that capture the full range of skills for those contexts and eliminate problems with ceiling effects. This dynamic perspective on developmental science starts with the variability of skills and finds the principles of order within that variability, which contrasts with the traditional view of abilities as fixed and static.

If psychology tells richer, more accurate stories, then its relation to policy needs to be reframed. Psychologists cannot draw simple lines in the sand, after which a developing person can be confidently assigned a full cognitive or psychosocial capability. Development is more complex and variable than that. Similar problems have surfaced in the emerging field of mind, brain, and education, where many people want to use neuroscience and genetics to shape and define education policy (Stein & Fischer, in press). Admitting the true complexity of human developmental processes means that, following Habermas (1996) and others, scientists and disciplinary experts should not claim that research results require a certain policy. Instead, they need to help inform the public debate about a policy by presenting relevant evidence that can illuminate the ultimately evaluative decisions about what policies to set, all perspectives considered and given the facts in hand. This approach is consistent with Steinberg et al.’s (2009) argument for “a careful and nuanced consideration of the particular demands placed on the individual for ‘adult-like’ maturity in different domains of functioning” (p. 593).
REFERENCES


