Abstract
This paper’s purpose is to provide a foundation for viewing critical thinking as both a maximal and typical performance construct. While maximal performance measures the best a person can do, typical performance measures what the person is most likely to do. An overview of maximal performance, including its history and limitations, will be given. The role of maximal and typical performance in cognitive development will be demonstrated through an exploration of the relationships between behavior, the environment, personality, crystallized intelligence, and fluid intelligence. Furthermore, these topics will be related to the development and use of critical thinking skills. Discussion will conclude with directions for future research.

Keywords: typical performance, cognitive ability, everyday performance, critical thinking

A Rationale for Considering Typical Critical Thinking Skills

School assessments commonly explore areas such as cognitive ability, academic achievement, personality, and behavior. Tests of cognitive ability and academic achievement are designed to elicit best performance (Furnham & Chamorro-Premuzic, 2004). For this reason, they are called maximal performance tests (Cronbach, 1949). Measures of personality and behavior focus on what an individual usually does; therefore, they are called typical performance tests (Cronbach, 1949). Although assessment measures are commonly divided into these categories (Reynolds, Livingston, & Wilson, 2006), consideration of both constructs may provide a more complete picture of the student than is possible through considering only one. As a specific function of cognitive ability, critical thinking utilizes both cognitive ability and personality (i.e., dispositional) factors (Halpern, 1998; Ku, 2009). As such, it is conceivable that an individual’s critical thinking skill is best understood by considering both maximal and typical performance.

Overview of Maximal Performance
Maximal performance is contingent upon capacity and ability. Capacity is an individual’s “hypothetical potentiality for” (Cronbach, 1949, p. 13) the trait being measured. Cronbach (1949) defined ability as a “person’s performance on a task at present, with maximum motivation [and ideal conditions,] but without further training” (p. 13). Ackerman (1994) argued, “it may not be possible, even in theory, to specify the conditions that would allow an individual to . . . approximate his/her theoretical capacity” (p. 4). Because of the difficulty in trying to measure capacity, maximal performance tests provide an estimate of ability.

The first major test of cognitive ability was developed by Binet and Simon in 1905 (Kaplan & Saccuzzo, 2001). It consisted of tasks such as recognizing food, defining common objects, and distinguishing between abstract terms, with the purpose of determining which children could benefit from schooling (Kaplan & Saccuzzo, 2001). Because of this measure’s success in predicting school achievement, a new age began in the assessment of cognitive ability. To this day predicting school achievement remains one of the main purposes of estimating intelligence.

Fiske and Butler (1963) cite two reasons why cognitive functioning has been assessed primarily through maximal performance measures. First, maximal performance measures are used to obtain a “pure measure” (p. 253) determined mainly by capacity, removed as much as possible from outside influences. Second, maximal performance in a controlled setting is considered more stable “than performance under more lifelike conditions” (p. 253). In order to produce optimal conditions to elicit an examinee’s best performance, manuals of cognitive ability tests instruct examiners to reduce external distracters, such as choosing a “well-lit room with adequate ventilation” (C. R. Reynolds & Kamphaus, 2003, p. 16). They also commonly instruct examiners to help reduce or maximize internal mechanisms that affect performance by building rapport, providing praise for effort, and telling the examinee to “do the very best you can” (p. 41).

Although maximal performance measures attempt to remove the effects of spurious external influences, the complete elimination of such influences is improbable (Kaufman & Lichtenberger, 1999). In addition, the influence of internal factors can never be fully removed from a test score. To consider the effect of common internal factors, such as depression, test manuals often report data on how the scores differ between individuals with and without these diagnoses (C. R. Reynolds & Kamphaus, 2003; Wechsler, 1997). Interpretation guides also provide
Reasons for Considering Typical Performance

(1) In contrast to maximal performance, typical performance estimates mean or modal performance to predict most likely real life behavior (Cronbach, 1949; Fiske & Butler, 1963). As Cronbach (1949) stated, “abilities and capacities define limits of performance, but what one actually does is rarely motivated to the point where he uses his utmost quantity or quality of performance” (p. 305). Whereas maximal performance tests measure behavior in a controlled setting, typical performance tests measure behavior under everyday situations and motivation (Ackerman, 1994; Cronbach, 1949; Fiske & Butler, 1963), thus providing further information about an individual that may be useful.

In addition, (2) typical performance may be better than maximal performance at predicting long term performance (Ackerman, 1986), and (3) at predicting performance in the later stages of skill acquisition (Ackerman, 1986, 1994; Goff & Ackerman, 1992). (4) Assessing one’s typical use of critical thinking skills may also provide additional information about strengths and weaknesses (Ackerman, 1994; Dennis, Sternberg, & Beatty, 2000). (5) Measuring typical performance may be more cost effective and efficient than measuring maximal performance (Dennis et al., 2000). (6) It may also help evaluate response to interventions (Diamond & Squires, 1993). These six reasons will be discussed in greater detail.

First, typical performance may be better than maximal performance at predicting how an individual will behave in real life. Because measures of typical performance measure mean or modal behavior, they may provide a better prediction of future performance on daily tasks than maximal performance measures (Ackerman, 1994; Goff & Ackerman, 1992). In studying problem solving of individuals with frontal lobe dysfunction, Channon (2004) argued “the dangers of ignoring real-life performance in favor of focusing only on controlled laboratory experiments...are substantial” (p. 238).

When considering typical performance, best possible performance is not as important as performance on a daily basis. Cronbach (1949) explained the value of assessing how an individual behaves in an everyday environment with the following example:

> There is little value in determining how courteous an applicant for employment in a store could be when she wanted to; almost anyone of normal upbringing has the ability to be polite. But the test of a suitable employee is whether she maintains that courtesy in her daily work, even when she is not especially motivated or “on her best behavior.” (p. 14)

This phenomenon is further exemplified by Stanovich’s (2009) examination of cases where an individual fails to “think or behave rationally despite having adequate intelligence” (p. 2) in what he termed “dysrationalia” (p. 2). On this basis, he argued that although higher cognitive ability can foster critical thinking, it does not guarantee that the individual will apply these critical thinking skills in daily life.

Second, typical performance may be better than maximal performance at predicting long term performance because of the effects of personality, motivation, and self-regulation. One example of the long-term predictive ability of a typical performance measure is grade point average (GPA) (Rolfhus & Ackerman, 1999). Halpin, Halpin, and Schaefer (1981) demonstrated that high school GPA was a better predictor of the GPA of college freshman than leading college entrance exams.

Third, typical performance measures may be better suited for assessing performance in the later stages of skill acquisition (Ackerman, 1994; Goff & Ackerman, 1992). Ackerman (1986) argued, “investigators conclude that individual differences in intellectual abilities poorly predict (i.e., r = .2 to .4) task performance as time and number of trials on the tasks increase” (p. 102). Contrasting the predictive qualities of typical behavior and maximal cognitive ability, Ackerman (1994) stated that maximal performance is associated with skill acquisition in its early stages or “the cognitive phase of skill acquisition” (p. 13). After this phase, the cognitive demands are reduced and other factors...
related to typical performance, such as personality and motivation, are more influential in determining the level of proficiency. As Ackerman (1994) stated “when there is a mismatch between intelligence-as-typical performance and intelligence-as-maximal performance, the learner may be satisfied with a suboptimal level of task performance, and essentially ‘exit’ form [sic] the learning process” (p. 13).

Fourth, assessing typical performance may provide a more comprehensive view of an individual’s strengths and weaknesses than is possible by looking only at maximal performance (Ackerman, 1994; Dennis et al., 2000). For example, through studying everyday behaviors, Smith-Park, Fawcett, Nicolson, and Fisk (2004) discovered that adults with dyslexia showed statistically significant impairment in organization, absentmindedness, and attention compared to adults not diagnosed with dyslexia. As another example, McDermott (1999) discovered that academically successful students were more likely to participate actively, listen attentively, and accept feedback than their less successful peers.

The discovery of strengths and weaknesses in typical performance relative to maximal performance may be particularly relevant. The discrepancy between typical and maximal performance may reveal a potential for improving performance (Gilbert, 1978). For example, an individual may be less susceptible to only consider self-confirming information or myside bias within a laboratory setting, but more susceptible in real life. Differences between maximal and typical strengths and weaknesses may occur for several reasons, including the individual’s personality (Goff & Ackerman, 1992), interests (Ackerman, 2003), motivation (Cronbach, 1949), or self-regulation (Ackerman, 1994).

Fifth, measuring typical academic and cognitive skills may be more cost effective and efficient than measuring maximal performance (Dennis et al., 2000). Typical performance is commonly measured via rating scales (Cronbach, 1949). Because of the relatively low cost and time requirements on the part of the assessor, a measure of typical performance could be used as a screening device to guide further testing (Diamond & Squires, 1993). For example, saving assessment time through adequate screening measures allows children who need specialized services to receive them with less delay.

Finally, consider the role of typical performance may help monitor response to intervention. Because typical performance measures of academic and cognitive skills relate more specifically to learning behaviors (Diamond & Squires, 1993; McDermott, 1999), the strengths and weaknesses found through these measures may easily translate into target behaviors for change (McDermott, 1984). Because of this focus, a measure of typical performance may also be able to detect subtle changes in behavior or cognitive processing that have not yet manifested themselves through maximal performance test scores.

Role of Everyday Behavior in Development of Cognitive Ability

One of the most compelling reasons why typical performance should be considered is the probable role of typical performance on cognitive development. The influence of typical behavior on cognitive performance and academic achievement is intertwined with several factors. The most prominent among these are genetic endowment (Jensen, 1973), personality (Ackerman, 1994), behavior (Flynn, 1998), and the environment (Dickens & Flynn, 2001). Bandura (1977) was one of the first authors to integrate these factors, in what he called “triadic reciprocal determinism” (Bandura, 1986, p. 23). This framework consisted of the person, the environment, and behavior, with influences such as genetic endowment and personality included in the person (Bandura, 1986). Bandura’s (1986) model emphasized a “mutual action between causal factors” (p. 23).

Similar to Bandura’s (1986) model, the model presented in Figure 1 represents the integration of research and arguments presented from divergent fields related to the nature of cognitive ability and achievement. To avoid an overly-cumbersome model, discussion of the development of cognitive ability will be limited to the relationships between five constructs: fluid intelligence (g_f), personality, behavior, the environment, and crystallized intelligence (g_c)/domain specific knowledge. Each of Bandura’s components is represented; however, the person of Bandura’s model is subdivided into three parts: personality, g_c, and g_f. The relationships between factors were also modified (See Figure 1).

![Figure 1](image)

Cattell (1966) defined g_c as “that form of general intelligence which is largely innate and which adapts itself to all kinds of material, regardless of previous experience with it” (p. 369). Personality is one’s attitudes, beliefs, and interests. Behavior was defined by Barker (2001) as “the way in which an animal acts or responds within the
environment” (p. 435). The environment includes any stimuli outside the person. The last concept, g, was defined by Cattell (1966) as “a general factor, largely in a type of abilities learned at school, representing the effect of past application of fluid intelligence, and amount and intensity of schooling” (p. 369). For the purposes of the current model, g, and achievement will be considered together; however, it should be noted that Cattell’s model of intelligence distinguishes between these concepts.

Innate ability and the environment are interconnected in determining cognitive ability (Neisser, 1998); however, the contribution of each factor has been heavily contested (Cattell, 1971; Dickens & Flynn, 2001; Flynn, 1998). Goldstein and Reynolds (1999) found that proposed ratios varied from 80:20 to 20:80. They stated that “even if 80% of an individual’s intellectual level is genetically determined, changes in intellectual level as a function of environmental influences and transaction may be enormous” (p. 5).

The environment and IQ are considered highly correlated (Flynn, 1998). This has been evidenced in the gradual, consistent rise in IQ since the invention of modern intelligence tests (Flynn, 1987). This phenomenon has been termed the Flynn Effect. As Neisser (1998) stated “the fact that (unknown) environmental factors are raising the mean IQ of Americans by 3 points per decade certainly shows that the environment matters!” (p. 15).

The environment’s role in determining cognitive ability becomes more complex when considering that individuals may shape their environment differently depending on innate ability (Jensen, 1973). Sternberg (1997), for example, argued that history’s great innovators are distinguished by their willingness and skill in shaping their environment. Marin and Halpern (2011) found that individuals with higher critical thinking skills were more likely to be enrolled in science courses than those with lower critical thinking skills. Several formulas have been spawned from the arguments regarding the relationship between innate capacity and the environment in determining ability (Dickens & Flynn, 2001; Turkheimer, 2004). The present model differs from these by considering the role of personality in determining behavior that, in turn, shapes the environment.

Much has been written on the role of personality in the development of academic and cognitive skills. Wittmann and Stiβ (1999) contended “knowledge is influenced by intelligence, personality, interests, and motivation” (p. 86). Ackerman (2003) argued “trait complexes help determine the direction and level of effort toward knowledge and skill acquisition” (p. 92). Cattell (1966) even contended that intelligence may be a personality trait. Typical cognitive abilities and personality traits are thought to be most closely related to g, and domain specific knowledge (Ackerman, 1994, 2003; Goff & Ackerman, 1992; Rolfus & Ackerman, 1999) because g, is thought to be a product of an individual’s experiences in the environment (Goff & Ackerman, 1992; Papalia, Olds, & Feldman, 1998).

Personality serves as a moderator between behavior and cognitive ability. This moderation is done through the behavior of selecting the environment. For example, a child with an interest in dinosaurs may choose to read about dinosaurs. Reading would increase the child’s knowledge of dinosaurs, and most likely would increase the child’s reading ability and vocabulary. In this example, personality affected the behavior, including the selection of the environment (i.e., exposure to reading). Over time, differences in behavior between two children with the same cognitive ability may account for a divergence of academic achievement and cognitive ability, and differences in domain specific knowledge.

The arrow in Figure 1 from environment to personality signifies that personality is also influential in interpreting information gathered through the environment (McCaul, 1944). Interests and temperament may influence the environmental information to which one attends (Driver, 2001). Personality is also involved in assigning meaning to events (e.g., joke verses criticism; Weiner, 1986). In this model, there is not a direct link from personality to environment to indicate the view that personality does not influence the environment directly, but does so through behavior. For example, upset babies receive comfort because they are crying, which indicates they are upset (i.e., the behavioral manifestation of the emotional state).

Personality may also account for differences in the time spent processing information, further leading to differences in g, (i.e., spend more time thinking about a topic based on interest). As Cattell (1966) contended “what we call crystallized intelligence [g,] is the collection of skilled judgements a person has acquired by applying his fluid intelligence to his school opportunities” (p. 304). For example, Rodriguez (2009) discovered individuals with higher academic self-concepts were more likely to use deep-learning strategies instead of strategic approaches. Conversely, the role of personality on information processing may also be influenced by differences in g, given that g, may influence the amount of effort and time required for gains in g,. Thus, the effort required for learning may influence one’s attitude toward the subject, further showing the interconnectedness of these processes.

To summarize, the development of cognitive ability is determined by the interactions of several factors. Learning and achievement are not only dependent on the abilities with which one is born, but also what one chooses to do with those abilities as a result of personality influencing behavior. Typical behavior, therefore, may be an important determinant in cognitive ability and academic achievement. Likewise, critical thinking skills and their utilization are dependent on cognitive ability, disposition, and prior learning (Halpern, 1998; Rodriguez, 2009; Stanovich, 2009).
Typical Cognitive Performance Within a Critical Thinking Context

Similar to the application of other cognitive and academic skills, an individual’s use of critical thinking skills may also vary by context. How one typically utilizes critical thinking skills may be more important than how those skills are utilized under optimal conditions. To explain this, Stanovich’s (2009) taxonomy of thinking errors will be utilized.

Stanovich (2009) contended that thinking errors can generally be categorized into two primary domains, problems due to the “cognitive miser” and “mindware problems.” The cognitive miser refers to the tendency of individuals to rely on problem-solving tactics that are quick and require less concentration over processes which “require great concentration that is often experienced as aversive” (p. 63). Mindware refers to “rules, knowledge, procedures, and strategies that a person can retrieve from memory in order to aid decision making and problem solving” (p. 129).

The concept of the cognitive miser easily fits with the prior discussion of typical critical performance. As mentioned previously, personality influences what cognitive resources are allotted in the processing and encoding of information (Ackerman, 2003). In Stanovich’s (2009) taxonomy, the cognitive miser goes further, explaining the various cognitive fallacies that can lead to less cognitive effort being utilized for optimal critical thinking. In this sense, the individual may fall into one of multiple cognitive traps such as myside bias (Stanovich & West, 2007), attribution substitution, or override failure (Toplak, Liu, Macpherson, Toneatto, & Stanovich, 2007). In contrast to Stanovich’s taxonomy, the model depicted in Figure 1 more explicitly emphasizes the role of the environment in the choice to utilize less precise heuristics. Because it is probably unwise to utilize one’s maximal ability to think critically in all situations, then the key factor may be to discern accurately how much cognitive effort is optimal in a particular setting.

In regard to mindware problems, Stanovich (2009) indicates that many difficulties are caused by “contaminated mindware” (p. 179; i.e., faulty problem-solving strategies) and “mindware gaps” (p. 179; i.e., deficient or non-existent problem-solving strategies). The mindware gap can be due to gaps in domain specific knowledge. For example, an individual lacking basic knowledge of cars and rules of negotiation would have difficulty selecting and purchasing an automobile (i.e., optimal critical thinking is impossible because the person lacks the needed tools to do so).

The concept of contaminated mindware provides an added layer from Stanovich’s (2009) taxonomy which is not covered in the model depicted in Figure 1. “Contaminated mindware” can be considered analogous to cognitive distortions. As Gannon (2009) explains, individuals develop their own working models or schemas of the world and even how to solve problems. These schemas can be based on false beliefs. As with Stanovich’s (2009) cognitive miser, it is possible that inherent differences in personality can account for susceptibility to developing contaminated mindware. In addition to personality differences, the environment may also play a role in developing these faulty strategies/beliefs. One example of contaminated mindware is superstition (Toplak et al., 2007). B. F. Skinner (1948) found that when pigeons were given food at regular intervals, operant conditioning occurred even when the delivery of food was not contingent on a behavioral response. It was observed that when a random behavior occurred before the delivery of food, the animal would associate that behavior with food delivery and repeat it. Similarly, each time a lay psychology belief appears to be confirmed by a random association, the strength of that belief will increase, further entrenching contaminated mindware.

Methods of Measuring Typical Critical Thinking

The value of considering typical critical thinking skills has few practical applications unless it can be measured. Measuring typical critical thinking skills requires some level of aggregation of behavior over time. This can be done either by the assessor completing a large number of behavioral observations or by enlisting the help of individuals who have already conducted such observations informally through daily interactions, such as parents, teachers, or the individual (Cronbach, 1949, 1984). Behavioral observations allow the assessor to see behaviors of interest firsthand. If the behavioral observations are structured and conducted by competent individuals, such observations would theoretically be the best indicator of typical performance (Cronbach, 1949); however, determining typical performance through observation can be very time consuming. In describing the difficulty in determining typical performance through behavioral observation, Cronbach (1949) stated:

It is doubtful if one ever has a truly typical day. Typical behavior could be described as an average or composite of many single behaviors....To observe typical behavior, one must in some way obtain a sample of all relevant situations and all times when the situation arises. (p. 305)

In summary, many observations would be required to know if the observed behaviors were typical (Cronbach, 1949, 1984). Additionally, to make generalizations from naturalistic observations, behaviors must be observed across different settings and times of day (Cronbach, 1984).

One way to work within this limitation would be to develop measures of typical critical thinking skills that are domain specific. Although the results may not be very
generalizable, they would be of value for individuals with particular interest in performance within that domain. One such example is in the field of engineering, where employees may not be worried about the application of critical thinking skills in daily life but rather how consistently individuals utilize such skills within their areas of expertise (Ralston, Larson, & Bays, in press).

A current longitudinal study (Ralston et al., in press) reports work on assessing the critical thinking skills of undergraduate engineering majors. Specially-trained faculty annually score specific writing samples using rubrics developed from the Richard Paul-Linda Elder critical thinking framework (Paul & Elder, 2006). Thus, those scoring the writing samples are utilizing a specific behavior (i.e., writing) within a specific field (i.e., engineering). The presented inter-rater reliability data suggests this measure may have some promise. By taking the scores obtained across each of the four years of participation, the authors may be able to provide a measure of the students’ typical use of or ability to express their critical thinking process within the engineering domain.

Rating scales are the most common method used for determining typical performance (Cronbach, 1949). Rating scales are commonly completed by individuals other than the assessor, reducing the assessor’s administration time, thereby reducing the cost of the assessment (Hart & Lahey, 1999; Saudino et al., 1998). Rating scales of typical academic and cognitive performance have utilized the ratings of teachers on their students (Hammill & Bryant, 1998), children on their elderly parents (Williams, Klein, Little, Haban, 1986), parents on their children (Dewey, Crawford, Creighton, & Sause, 2000; Dewey, Crawford, & Kaplan, 2003; Williams et al., 1986; Williams, Ochs, Williams, & Mulhern, 1991), and students rating their own behavior (Kruger & Dunning, 1999). For critical thinking, the California Critical Thinking Disposition Inventory (CCTDI) has been used to assess critical thinking dispositions in nurses in the United States (Zori, Nosek, & Musil, 2010), China (Hwang, Yen, Lee, Huang, & Tseng, 2010), and Korea (Ju An & Sook Yoo, 2008).

Rating scales have been preferred over behavioral observations for measuring typical academic and cognitive performance (Dewey et al., 2000; Dewey et al., 2003; Williams et al., 1991; Yen, Konold, & McDermott, 2004). Kratochwill, Sheridan, Carlson, and Lasecki (1999) cited multiple strengths of rating scales and checklists over observation and interviewing alone. These strengths include cost, time, and effort efficiency; a greater likelihood of obtaining a comprehensive view of the construct of interest; and easily quantifiable information. Kratochwill et al. further explained that information in this format (relative to behavioral observation or interview data) can be more readily used to assess pre and post treatment performance. Furthermore, because rating scales of typical performance can be more easily administered to large groups, information can be obtained to identify individuals who may respond well to a particular intervention aimed at improving critical thinking skills.

Another reason for the use of rating scales is that it reduces the limitations of behavioral sampling found in ability testing. This is because the rater mentally averages the behaviors observed during a specified time frame (e.g., past two months) to provide an overall estimate of typical behavior (Anastasi & Urbina, 1997; Glascoe & Dworkin, 1995; Kratochwill et al., 1999; Saudino et al., 1998). Because of this compilation of observations, the results may provide a closer approximation to what the individual is most likely to do in an everyday setting than a test of maximal performance can provide.

The utility of rating scales becomes more apparent when considering real-world applications of their use. One such example is in the realm of psychological assessment. Because assessments are individually tailored to each examinee, it is impossible for training programs to provide rigid algorithms for testing decisions. As a result, being able to assess the student’s decision-making process across the various practicum experiences can help trainers understand how well students will be able to act independently after graduation. Similarly, rating scales of typical critical thinking skills may be useful within the medical field, where physicians must find ways to diagnose and treat various illnesses while avoiding excessive testing.

Because rating scales depend on a rater estimating behavior, this form of measurement is not as accurate as compiling data through many direct observation sessions. Many factors affect the accuracy of rating scales, and here are some guidelines for dealing with two major issues. First, ratings are depending on the rater’s understanding of the directions and what is being asked by each item (Cronbach, 1949; Kratochwill et al., 1999). Second, the subjective estimation is limited to the exposure the rater has had with the subject in the areas targeted on the rating scale (Cronbach, 1949, 1984; Glascoe, 2001). Several steps can be taken to reduce the possibility that confusion about item content will influence the ratings. Expert raters can be enlisted to help determine potentially confusing or culturally offensive items (Anastasi & Urbina, 1997). The ratings for and by individuals from different identifiable groups of people can be assessed to determine if consistent differences between these groups exist, called differential item functioning (Kaplan & Saccuzzo, 2001). The ratings of items can also be used to determine which items detract from the internal consistency of the scale as a whole. This is done with the assumption that less reliable items may be the result of different views of what the item is asking, or that the item relates to a less-related aspect of the construct (Kaplan & Saccuzzo, 2001). Cronbach (1984) gave five guidelines for item content in a rating scale. “Each rating should refer to a single variable” (p. 514). “There should be a number of items touching on the same aspect of behavior” (p. 514). “Scales should describe the strength of a trait...and not present ‘opposite’ traits as a bipolar scale.”
"Items should be as free as possible of theoretical preconceptions" (p. 515). "The span of the scale should not extend beyond the range of cases" (p. 515). Glascoe and Dworkin (1995) stated that a valid scale depends upon the items matching the constructs being studied.

With regard to a second issue, it is often advantageous to obtain ratings from multiple raters (Glascoe, 1991; Reynolds & Kamphaus, 2004). One advantage is the ability to compare the agreement between raters. Another advantage is increased accuracy because of the greater likelihood of covering all aspects of development (Glascoe, 1991).

**Future Directions for Research**

Research in the area of typical critical thinking skills could be developed in several areas. One of the more clear areas of development would be determining the effects of critical thinking skills in job performance in various settings, similar to the work that has been done within the nursing profession (Zori, Nosek, & Musil, 2010). Further work in this area could prove useful for direct application in the selection of employees for hire or promotion. Given the relatively small number of studies found attempting to assess critical thinking skill via rating scale, future research may indicate differences in accuracy based on specific skills or fallacies assessed. Similar to the study of typical cognitive engagement, further exploration of the power of typical performance measures to predict performance in the later stages of skill acquisition (Ackerman, 1994; Cronbach, 1949; Goff & Ackerman, 1992) may provide new insights to the effects of critical thinking over time.

Research also could focus on the discrepancy between typical and maximal performance. Although it has been demonstrated that direct teaching can have lasting effects on the usage of critical thinking skills (Riesenmy, Mitchell, Hudgins, & Ebel, 1991), it is possible that, similar to other performance tasks, one would not always utilize critical thinking ability to the fullest extent. Similar to the finding that perceived organizational support increases maximal work behavior within the job site (Witt & Spitzmüller, 2007), further investigation of factors leading to utilizing higher order thinking skills may lead to many real world applications.

**Summary**

Tests of maximal performance measure best performance, while tests of typical performance measure average performance. Maximal performance tests are measures of ability and not capacity. Because maximal performance tests measure ability, the scores on maximal performance tests can fluctuate over time beyond measurement error alone. The measurement of typical academic and cognitive behavior, such as critical thinking skills, has many benefits, including prediction of real-life behavior; prediction of long-term performance; providing additional information about strengths and weaknesses; and being more cost effective and efficient than measuring maximal performance. Finally, typical behavior may play an important role in the development of cognitive abilities over time.

**References**


**Author Information**

Gordon D. Lamb, Ph.D. (Texas A&M, 2008) is an Assistant Professor at Sam Houston State University. He teaches courses on child assessment, learning, and development. His research has covered several topics related to psychological assessment including autism assessment, health-related quality of life, and everyday academic and cognitive skills. Correspondence should be addressed to: Gordon Lamb, Department of Psychology & Philosophy, Sam Houston State University, Campus Box 2447, Huntsville, TX 77341. Email: gd003@shsu.edu.

Cecil R. Reynolds, Ph.D. (U. of Ga., 1978) is the author of more than 300 scholarly publications and author or editor of more than 50 books including *Handbook of School Psychology*, the *Encyclopedia of Special Education*, and the *Handbook of Clinical Child Neuropsychology*. He is also the author of several widely used tests of personality and behavior including the Behavior Assessment System for Children-second edition, Reynolds Intellectual Assessment Scales, the School Motivation and Learning Strategies Inventory, the Koppitz-2 Developmental Scoring System for the Bender-Gestalt Test, and the Revised Children’s manifest Anxiety Scale. Dr. Reynolds is a past president of APA Divisions 5, 16, and 40. He is currently editor of Psychological Assessment, and a former editor of Applied Neuropsychology and Associate Editor of School Psychology Quarterly, and serves as an Emeritus Professor of Educational Psychology, Professor of Neuroscience, and Distinguished Research Scholar at Texas A&M University.