# DEDUCTIVE REASONING: THE USE OF THE SELECTION TASK TO MEASURE EDUCATIONAL INTERVENTIONS

***John D Rich, Jr., PhD***
Delaware State University/Psychology Department,
United States of America

**Abstract**

A common problem with educational interventions is that they require the collection of data across one or many school years before an assessment of their effectiveness can be performed. This paper describes the selection task, and makes the case for its use in evaluating the effectiveness of educational interventions. As an example, a study is described that uses the selection task to determine the degree to which student performance is positively influenced by knowledge of success criteria. Further, this study examined whether the impact of knowledge of this success criteria is inhibited or augmented by the magnitude of the success criteria provided. Participants were given success criteria of 10/10, 8/10, 6/10, 4/10, or nothing (control). Performance on the task was greatest when success criteria were provided, but only when the success criteria were not too strict. Implications for educational practice are discussed.

**Keywords:** Deductive Reasoning, Selection Task, Education

## Introduction

The specific aim of the present study is to address whether or not providing knowledge of criteria facilitates success on an academic task. In addition, this study will seek to replicate findings on the differential difficulties between two types of problems on a commonly used measure of deductive reasoning. The impact of teacher expectations on student performance has strong research support. As students perceive higher expectations from their teachers and/or parents, academic performance tends to increase. (Babad, 1998; Brophy, 1983; Cooper, 1985; Feldman and Thiess, 1982; Good and Weinstein, 1986; Haynes, Tikly, and Caballero, 2006; Rubie-Davies, 2006; Rubie-Davies, 2007; Rubie-Davies, Peterson, Irving, Widdowson, and Dixon, 2010; Weinstein, 2002)

However, these studies have focused on increasing teacher expectations among student populations who had previously perceived teacher expectations to be significantly lower. If expectations are too high, students may be inclined to decrease effort in order to avoid failure. (Grant and Dweck, 2003; Hong, Chiu, Dweck, Lin and Wan, 1999; McWhirter, Torres, Salgado, and Valdez, 2007; Plaks, Grant and Dweck, 2005).

According to the attribution theory developed by Bernard Wiener (1980, 1982), motivation is somewhat like a cost-benefit analysis that weighs effort and risk against the expectations for success. Working hard and doing well can provide a real joy and pride, but one may not feel like the odds of this happening are worth the risk. If one really tries, and he or she do not succeed, then he or she is almost forced to conclude that one is what one fears – stupid. Putting forth little or no effort offers one the opportunity to brush off a failure by attributing the failure, not to one's lack of ability, but rather to one's lack of effort. Over time, students who have tried and failed may exhibit an educational form of learned helplessness, where failures are expected and successes are attributed to external causes (such as luck). (Dweck, 1975)

When a student is given specific guidance, such as a rubric, which clarifies criteria for grading success, the result can be an improvement in performance. Some studies indicate that rubrics clarify expectations, and make grading standards more comprehensible (Bissell and Lemons, 2006; Dahm et al., 2004; Frederiksen and Collins, 1989). There is, however, a dearth of literature on the actual effectiveness of rubric use on student's academic performance. Rubrics distinctively set standards for achievement; they also provide a specific way to grade work and provide students with helpful feedback they can later revise. Rubrics can possibly make expectations clearer to students, especially when a rubric is used as a guide while completing an assessment (Allen and Tanner, 2006; Arter and McTighe, 2001; Dochy, Gijbels, and Segers, 2006; Perlman, 2003). It seems that students are asking for guidance and a clear vision of what the teacher desires, helping them to better understand the goals for their project. When a professor uses rubrics to outline what information they want answered, students can easily narrow down and eliminate other information to perform better on an assessment.

In one experiment (Howell, 2011), the control group who completed a course assignment without a grading rubric, scored, on average, considerably lower, than the treatment group, who completed the same assignment, but was provided with a grading rubric. This research supports the idea that students who are provided with a grading rubric, as compared to their colleagues who are not given a rubric, perform better on an assignment (see also Moskal, 2000). The use of rubrics as an instructional tool goes hand and hand with the clarity of a student's expectations and the student's ability to set realistic goals to achieve. Researchers argue that rubrics are beneficial to a students' peer and self-assessment (Goodrich, 1997). In this case a student is more motivated and confident in the work they have done when guided by the rubric. Students learn more on a concept when using the rubric as a guide (Andrade, 2000).

Not only can rubrics point out the strengths and weaknesses of student writing skills, they provide an outline of information needed so students can directly absorb the information needed for the assessment. A study by Andrade and Saddler (2004) concluded that if a category on a rubric was sentence structure, and students were encouraged to begin sentences in different ways, a student could assess his or her sentences based on the rubric and revise the sentences to meet the standards specified. Additionally, rubrics save time for teachers with multiple classes and students. Instead of commenting on every assignment due, the professor now can point out what is directly missing from the assignment on the rubric. A clear rubric can help students understand any distinctions or alignments between what was submitted and what the professor wanted. This information can then be used to improve the quality of work on future submissions.

In addition to providing information about criteria, it is also important that the criteria be within the means of each student. An academic task with criteria that is far above a students' level of competence will be ineffective. Vygotsky's concept of scaffolding suggests that knowledge must be appropriate for the child's level of comprehension (Eddy, 2010). In his zone of proximal development (ZPD) theory there are two levels of attainment for the ZPD: Level 1 represents the 'present level of development'. This describes what the child is capable of doing without any help from others. Level 2 represents the 'potential level of development'. This refers to what the child could potentially be capable of demonstrating with help from other people or 'teachers'. However, anything that is too complicated for the child to learn that isn't in their ZPD cannot be learned at all until there is a shift in the ZPD. When a child does attain their potential, this shift occurs and the child can continue learning more complex, higher level material.

Therefore, if the expectations for a student are realistic and reasonable, then an academic goal becomes achievable and more likely to be accomplished. According to Huba and Freed, (2000) a necessary feature for measuring student learning outcomes is to set benchmarks for successful student, course, program, or degree outcomes, including milestones to gradually move from current performance levels to the benchmark goal. This suggestion is in alignment with Vygotsky's theory of proximal development and is the guiding theory of the present study.

Two previously studied areas of research on the development of reasoning constitute the background for the construction of the present study. The first area is in the field of deductive reasoning, which particularly questions the role of logic versus the role of semantic content in determining performance on reasoning problems. One theoretical approach (O'Brien and Overton, 1982; Byrnes and Overton, 1986; Overton, 1990; Overton, 2010) maintains that the development of a logical competence is a necessary prerequisite for solutions to deductive reasoning problems. According to this approach, specific content effects operate as moderators of logical competence to inhibit or enhance successful task performance. From this perspective, both formal reasoning processes and the real-time information-processing strategies related to content effects play essential roles in determining task performance. A second theoretical approach (e.g., pragmatic schemas approach of Cheng and Holyoak, 1985; mental models approaches of Johnson-Laird, 2006 and Markovits and Barrouillet, 2002; metacognitive approaches of Kuhn and Franklin, 2006 and Moshman, 2004) asserts that formal reasoning processes play little or no role in task performance. According to this approach, task performance is virtually totally determined by context variables such as those provided by specific semantic contents.

This study sought to explore both the role of the development of formal logical reasoning and the role of knowledge of success criteria in the solution of deductive reasoning problems. The development of formal logical reasoning was represented in Piaget's theory (Inhelder and Piaget, 1958) as a succession of increasingly powerful logical systems or competencies. Several studies (Byrnes and Overton, 1986; O'Brien and Overton, 1982; Overton, Byrnes, and O'Brien, 1985) have supported the prediction that a fully developed formal logical competence is not available before adolescence. These studies, which focused specifically on conditional inferences, were not, however, solely concerned with the acquisition of logical competence. In addition, the research explored several factors, including task interpretation, cognitive style, and semantic content, as moderators of the logical competence once that competence had been acquired. This research is also framed by the "Dual systems Competence < --> Procedural processing" approach (Ricco and Overton, 2011).

This dual systems approach is a formulation, within a developmental perspective, of an integration of stage-structure (levels of competence) and information-processing (moderators) views of cognition. This approach maintains that performance on any cognitive task is a function of both level of logical competence and the individual's information-processing skills and capabilities. With respect to the present research, the main feature of this approach is that it distinguishes between the logical knowledge necessary for task solution (which individuals may or may not have available) and procedural informational processes (e.g., attention, encoding, representation, retrieval) that may facilitate or inhibit the accessibility of available logical knowledge.

Wason's (1966, 1983) selection task provides an appropriate domain for exploring issues concerning levels of logical reasoning competence and the effect of semantic content as the latter is mediated through informational processes. The standard, or abstract, form of this task presents participants with an array of four cards showing, for example, respectively, the letter E, the letter K, the number 4, and the number 7. In addition, participants are presented with a conditional rule: "If there is a vowel on one side of the card, then there is an even number on the other side." Participants are told that each card has a letter on one side and a number on the other and that the task is to select those cards and only those cards that would have to be turned over to decide with certainty whether the rule is true or false. A fully adequate solution consists of the selection of the E and the 7 card and only these cards. This solution recognizes that only a falsification strategy will provide logical certainty. The four showing cards may be described as the E = p, K = ~p (not p), 4 ~ q, the 7 - ~q (not q). The conditional rule "If p then q" is determined with certainty only under the condition in which it is falsified, that is, the p \ not q selection. Only the E card (p, possible not q on the opposite side) and the 7 card (not q, possible p on the opposite side) could together satisfy this condition.

A more common measure used for investigating developmental progressions in deductive reasoning is the adapted version of the selection task, which couches the conditional rule in more familiar, everyday scenarios (e.g. “If you have finished your homework, then you can watch TV.”) (Overton, 1990; Muller, Overton and Reene, 2001). In the version most commonly used, participants are given a rule in the form of “If p, then q,” along with four cards representing the affirmation and negation of both “p” and “q.” Participants are asked to demonstrate their ability to understand the rule as an implication, and determine the degree to which each of the four cards is permissible or not permissible, given the rule. Specifically, given “If p, then q,” then the combination “p and q” is permitted, “p and not q” is not permitted, and “not p and q” and “not p and not q” are indeterminate (for a fuller explanation of the selection task and its use in investigations of deductive reasoning, see Ricco, 2010).

Muller, Overton and Reene (2001) discovered that the selection task contains two types of problems. The first is the Action-Condition (AC) schema, where the consequent-clause is a precondition for the action or event mentioned in the antecedent. Thus, an AC problem states a relation of conditionality and can be translated as “*Q* is the necessary condition for *P*” (see Matalon, 1962; Matalon,1990) or "*P* only if *Q*" (Thompson, 1995). The explicit statement of the precondition relation highlights the asymmetric nature of the relation between antecedent and consequent; satisfying *Q* is only a necessary, but not sufficient condition, for the occurrence of *P*.

The second type of selection task problem is the Condition-Action (CA) schema. In this schema the particular temporal relation between antecedent and consequent does not suggest an asymmetric relation between antecedent and consequent; adults are also unlikely to rephrase an obligation as "*P* only if *Q*" (Thompson, 1995, Experiment 1). Thus, CA schema problems require the construction of an asymmetric relation between antecedent and consequent by negating the consequent and reversing the direction of the if-then sentence ("if *P* then *Q*; but not all *Q* are *P'*). Because the construction of an asymmetrical relation between antecedent and consequent involves complex operations and is cognitively demanding (Staudenmayer and Bourne, 1977), the *P* and *Q* cards are frequently selected in those selection problems (e.g., halls and school suspension) that do not make the asymmetric nature of the relation explicit. Muller, Overton and Reene (2001) found that CA schema problems are more difficult that AC schema problems. The purpose of this study was to replicate this analysis of selection task responses with a new sample and to examine the differential effect of the experimental manipulation on AC/CA schema problem success.

The ability to produce alternate antecedents that are meaningfully related to the consequent has been identified as a prerequisite for the correct solution to the Affirmation of the Consequent invalid argument form (Markovits, 1993). In the selection task, this solution entails not choosing the *Q* card. For example, to not choose the *Q* card in the drinking age problem, alternate antecedents (e.g., drinking Coke) must be generated and related to the consequent. Two explanations single out specific factors related to the production of alternate antecedents which account for the relative ease of choosing the *Q* card in AC problems as compared to CA problems. According to the first explanation (Barouillet and Lecas, 1998), the production of alternate antecedents is more difficult when the term in the consequent of the rule is binary (e.g., punished vs. not punished). This is based on the assumption that the initial representation of the conditional rule "*P* then *Q*" consists of the construction of a mental model that establishes a directional relation between *P* and *Q*. A binary consequent, thus, links *P* (e.g., running in the school halls) and *Q* (e.g., punished) and alternate instances of *P* (e.g., walking in the school halls) with *not*-*Q* (e.g., not punished). Accordingly, alternative instances of *P* are not linked to *Q*, which in the selection task would result in the selection of the *Q* card (i.e., the invalid form, Affirmation of Consequent). When the consequent is nonbinary, however, *P* is not linked to any determinate value of *Q*. As a consequence, links between alternative cases of *P* and *Q* (e.g., drinking Coke and being over 21 years of age) become possible, provided working memory allows the construction of more complex alternate models (for a detailed discussion, see Barouillet and Lecas, 1998; for evidence of developmental progressions in the use of argument forms, see Overton, Ward, Noveck, Black and O'Brien, 1987, Chappell and Overton, 1998, and Rich, Fullard, and Overton, 2011).

A second explanation for the relative ease of not choosing the *Q* card in AC as compared to CA problems, concerns the semantic relation between antecedent and consequent. A major problem in solving conditional reasoning problems consists of the ability to realize that an asymmetric relation holds between antecedent and consequent (i.e, if *P* then *Q*, but if *Q* then not necessarily *P*; see Staudenmayer and Bourne, 1977). Given the performance-competence distinction frequently discussed by researchers in this area, and given the theoretical underpinning of this study – that is, that college-aged students have the competence to correctly solve deductive reasoning problems, but do not always express this competence – another aim of this experiment is to continue attempts to facilitate success on the selection task through the examination of an independent variable hypothesized to increase student potential.

It is believed that these theories and findings about expectations can be applied to students' own self-motivations to achieve. In particular, the goal is to investigate whether knowledge of success criteria can improve or inhibit performance on a deductive reasoning task. Success on the selection task has been traditionally defined as a correct falsification on 6 out of 10 of the problems. In one of the criteria conditions, participants will be told that 6 out of 10 is the goal for successfully solving the task. It is believed that knowing that there is room for error, but also having a specific, realistic goal, will motivate students to try to solve the problems to a greater degree than in the control condition (no criteria provided), the low criteria condition (4 out of 10) or the high criteria conditions (8 out of 10 and 10 out of 10). In the low criteria condition, a hypothesis was made, favoring decreased rates of success, consistent with research on the dampening impact of low expectations discussed in the research above. In the high criteria conditions, it was expected for student performance to decrease in order to avoid possible failure, consistent with the findings of attribution theorists. Therefore, the general pattern of results which were expected is an inverted U-shaped pattern, with the highest rates of success in the 6 out of 10 (realistic) condition, and lower rates of success for the other conditions.

**I**

This study investigated the impact of progressively increasing criteria on deductive reasoning performance through the random assignment of one hundred fifty-nine African American participants (87 female, 72 male) to one of five conditions; namely, knowledge of different success criteria (control 4 out of 10, 6 out of 10, 8 out of 10, and 10 out of 10). The study used students from a sophomore-level psychology class that was limited to psychology majors.

Participants met in a large lecture hall, where they were seated far enough apart to not be able to view the information/paperwork of the participants on either side. Students were offered extra credit for their participation. All student responses were confidentially recorded.

**Instrument and Procedure**

Each participant received a test booklet. Test booklets contained an instruction page and ten selection task problems (“If *P* then *Q*”), each of which was presented on a separate page. The ten selection task problems are presented in Table 1; the problems are identical to those used in a study by Ward and Overton (1990) and Rich, Fullard, and Overton (2011). Six problems (drinking age, motor vehicle, homework, handgun, permission, and AIDS) were AC schemas in which the consequent specifies a precondition that must be met if the action specified in the antecedent is to be taken. Three problems (halls, school suspension, and murder) were CA schemas in which the consequent specifies an action that must be taken when the condition specified in the antecedent occur. The status of the tenth problem (retired) was more ambiguous because the antecedent “If a person is retired from work” can be interpreted as either specifying a precondition (i.e., given a person is retired, then the person is older than 55 years of age) or a consequence (i.e., given a person is over 55 years of age, then the person is retired) of the consequent clause “then the person is over 55.” The instruction page informed the participant that the booklet contained several problems and that each presented a rule. Participants were further informed that for each problem, they were to determine whether the rule was being broken. The experimenter read aloud the instructions while the participant read along silently. Using a practice problem about a public swimming pool as an example, the format of each page was as follows: At the top of each page were four response alternatives presented in rectangular boxes. Two of the response alternatives referred to the antecedent (the “if clause”), and two referred to the consequent (“then clause”) of the conditional rule. Two cards negated either the antecedent (*not-P*; i.e., sunbathing) or consequent (*not-Q*; i.e., no lifeguard), and two cards affirmed the antecedent (*P*; i.e., swimming) or consequent (*Q*; i.e., lifeguard). Below these alternatives were the following instructions:

Each of the above cards has information about four different people at the public pool. On one side of each card shows the person's behavior at the pool. One side of the card shows whether or not a lifeguard is present. Remember, you can only see one side of each card. Here is a rule: "IF A PERSON IS SWIMMING IN THE PUBLIC POOL, THEN A LIFEGUARD IS PRESENT". Select the card or cards that you would definitely need to turn over to decide whether or not the rule is being broken.

Table 1. Selection Task Problems

1. If a person is drinking beer, then the person must be over 21
   Cards: *P* = drinking beer; *not-P* = drinking Coke, *Q* = 22 years of age, *not-Q* = 16 years of age
2. If a student is watching television, then the student's homework is finished
   Cards: *P* = watching television; *not-P* = not watching television; *Q* = homework finished; *not-Q* = homework not finished
3. If a person is driving a motor vehicle, then the person must be over 16 years of age
   Cards: *P* = driving, *not-P* = not driving; *Q* = 18 years of age; *not-Q* = 14 years of age
4. If a student is caught running in the school halls, then the student must be punished
   Cards: *P* = running in school halls; *not-P* = not running; *Q* = punished; *not-Q* = not punished
5. If a student strikes a teacher in school, then the student is suspended from school
   Cards: *P* = striking teacher; *not-P* = not striking teacher; *Q* = suspended; *not-Q* = not suspended
6. If a person has a handgun, then the handgun must be registered
   Cards: *P* = handgun; *not-P* = knife; *Q* = registered; *not-Q* = not registered
7. If a girl under 14 years old has an abortion, then she must have permission from her parents
   Cards: *P* = 12 years of age; *not-P* = 16 years of age; *Q* = permission; *not-Q* = no permission
8. If a child with AIDS attends school, then that child has received approval from the community
   Cards: *P* = AIDS; *not-P* = sore throat; *Q* = approval; *not-Q* = no approval
9. If a drunken driver kills someone, then the driver must be charged with murder
   Cards: *P* = drunk; *not-P* = not drunk; *Q* = charged with murder; *not-Q* = not charged with murder
10. If a person is retired from work, then the person must be over 55 years of age
    Cards: *P* = retired; *not-P* = not retired; *Q* = 60 years of age; *not-Q* = 40 years of age

Space was provided below the boxes for the participant to mark his or her choice (or choices). The order of response alternatives was randomized across problems, and the problem order was randomized across participants.

**Scoring**

The logically correct answer for each problem consists of selecting the cards *P* and *not-Q* and not selecting the cards *not-P* and *Q*. The reason for this, as stated earlier, is that the *P* card and the not -Q cards, and only these two cards, could possibly each yield *P* and *not-Q* falsification instances. This selection combination, termed the complete falsification solution was awarded a score of 1. If this was not the response pattern given to a problem, a score of 0 was awarded. Performance on different argument forms (Modus Ponens, Denial of Antecedent, Affirmation of Consequent, and Modus Tollens) was scored separately. Correct answers to each argument form were awarded a score of 1, and incorrect answers were awarded a score of 0. The correct answer to Modus Ponens consists of selecting the *P* card, the correct answer to Denial of Antecedent consists of not selecting the *not-P* card, the correct answer to Affirmation of Consequent consists of not selecting the *Q* card, and the correct answer to Modus Tollens consists of selecting the *not-Q* card. For example, a participant who selected the cards *P* and *Q* would be awarded scores of 1 for the argument forms Modus Ponens and Denial of Antecedent, and scores of 0 for the argument forms Affirmation of Consequent and Modus Tollens. Finally, each problem was scored for the specific combinations of response alternatives selected by the participants. There were 16 possible selection combinations that could be given on each problem. One-way ANOVAs were calculated to examine the differential scores or response patterns between participants in the five experimental conditions.

**Results**

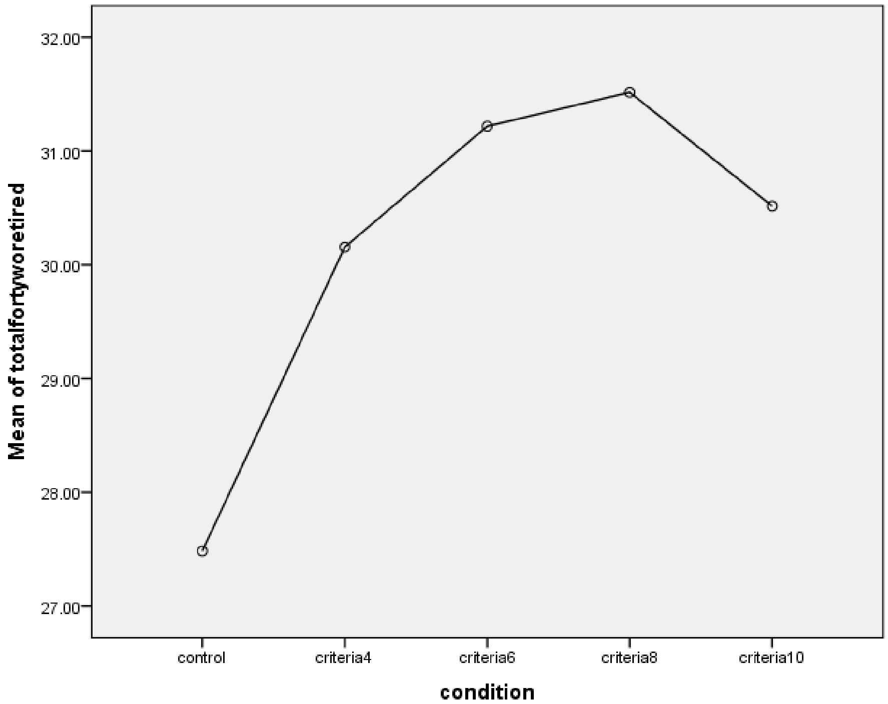
The percent of participants who answered each problem with complete falsification was first analyzed. These percentages were divided into AC and CA problems, consistent with Muller, Overton and Reene (2001). Inspection of table 2 indicates that the AC problems were easier than CA problems and that the ambiguous reasoning (retired) problem was easier than AC problems but more difficult than CA problems. Because there was interest in comparing performance on clear AC and CA problems, the ambiguous problem was dropped from further analysis.

| Problem | % of People had Problem Correct | |
|---|---|---|
| | AC | CA |
| Drinking | 89.3% | |
| Homework | 70.4% | |
| Driving | 84.9% | |
| Halls | | 28.3% |
| Retired | 51.6% | |
| Suspended | | 34.0% |
| Handgun | 74.8% | |
| Permission | 76.7% | |
| AIDS | 73.0% | |
| Murder | | 34.0% |
| Average | 76.2% (not including retired problem) | 32.1% |

In a second set of analyses, three dependent variables were assessed:

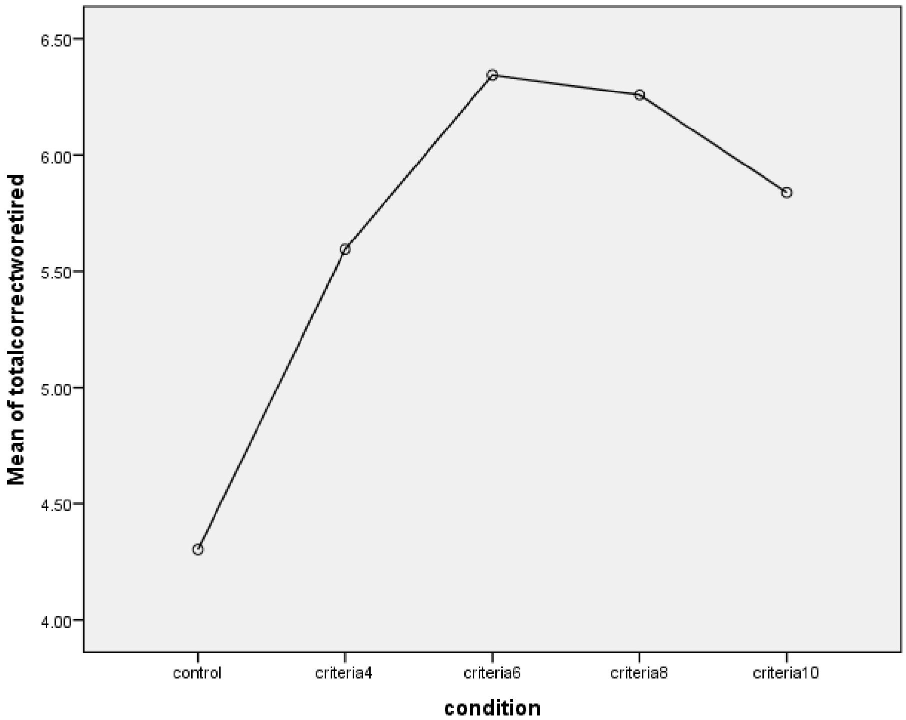
1. For the first dependent variable each of the 9 (the retired problem was excluded) problems could receive a score ranging from 0-4 (1 point each for choosing the p card, choosing the not q card, not choosing the not p card, and not choosing the q card. This yielded a potential range of scores of 0-36 for each participant. A one-way ANOVA, testing for significant differences between the control group and each of the four criteria groups, found a significant main effect: $F(4,154) = 4.860$, $p = .001$. Posthoc analyses revealed that scores for the control group (M 27.48, SD 5.06) were significantly lower than scores for the 6/10 group (M 31.22, SD 3.42, $p=.004$), the 8/10 group (M 31.52, SD 2.99, $p = .001$) and the 10/10 group (M 30.52, SD 4.05, $p=.039$). None of the criteria groups were significantly different from the others. In order to test the notion that providing participants with information on *any* criterion would facilitate significant improvements in performance over participants who were not given any information about the criterion, a separate ANOVA was calculated. This ANOVA revealed a significant main effect: $F(1,157) = 17.377$, $p<.000$. Figure 1 below shows the differences among the five experimental groups.

Figure 1. Complete Falsification Scores by Experimental Condition



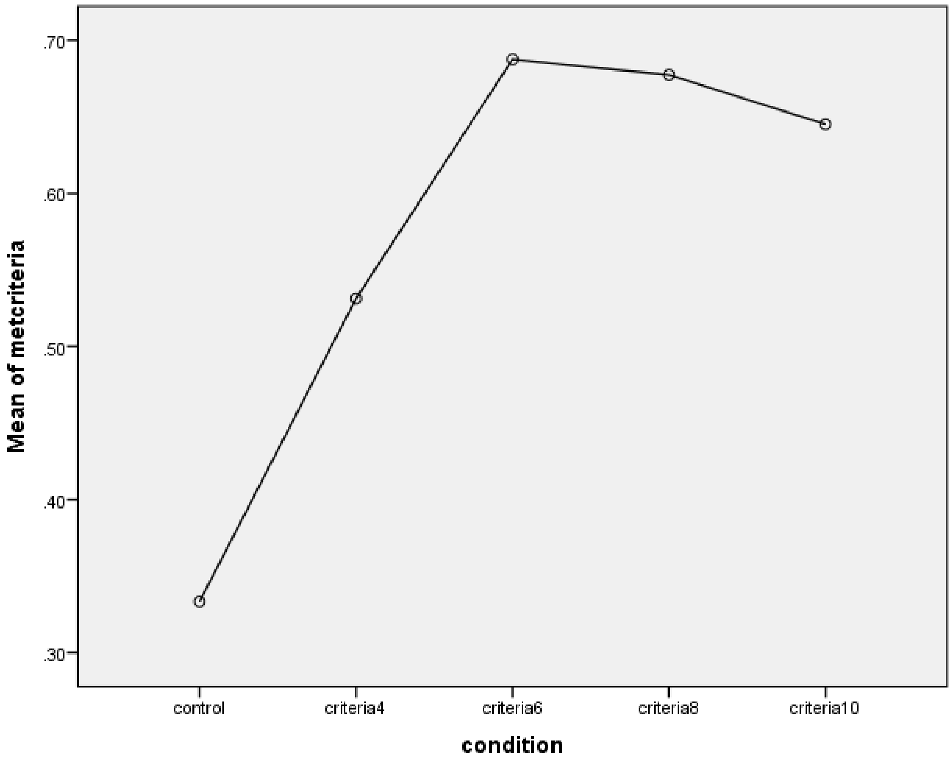
2. For the second dependent variable, each of the 9 problems (retired problem excluded) received a pass = 1, fail = 0 score. On each problem a participant was awarded one point for answering with a complete falsification (selecting the p card, not selecting the not-p card, not selecting the q card, and selecting the not-q card). Hence each participant's score ranged from 0-9. A one-way ANOVA, testing for significant differences between the control group and each of the four criteria groups, found a significant main effect: F (4,154) = 4.550, p = .002. Posthoc analyses revealed that scores for the control group (M 4.30, SD 2.51) were significantly lower than scores for the 6/10 group (M 6.34, SD 1.84, p=.003), and the 8/10 group (M 6.26, SD 1.91, p = .005). None of the criterion groups were significantly different from the others. In order to test the notion that providing participants with information on *any* criterion elicited significant improvements in performance over participants who were given no information about the criterion, a separate ANOVA was calculated. This ANOVA revealed a significant main effect: F (1,157) =15.787, p<.000. Figure 2 below shows the differences among the five experimental groups.

Figure 2. Dichotomous Complete Falsification Scores by Experimental Condition



3. The third dependent variable was similar to the second except that each participant received a score of 1 if the participant provided a complete falsification solution to at least 6 of the 10 problems. A one-way ANOVA, testing for significant differences between the control group and each of the four criteria groups, found a significant effect: F (4,154) = 3.065, p = .018. Posthoc analyses revealed that scores for the control group (M .33, SD .48) were significantly lower than scores for the 6/10 group (M .69, SD .47, p=.037), and the 8/10 group (M .68, SD .48, p = .050). None of the criterion groups were significantly different from the others. In order to test the notion that providing participants with information on *any* criterion elicited significant improvements in performance over participants who were not given any information about the criterion, a separate ANOVA was calculated. This ANOVA revealed a significant effect: F (1,157) =10.220, p=002. Figure 3 below provides a picture of the differences among the five experimental groups.

Figure 3. Percentages of Participants Meeting the 6/10 Success Criterion by Experimental Condition



These three set of analyses confirm the hypothesis of an inverted-U shaped pattern. Compared to the control group, the criteria groups scored significantly higher, with the criteria groups in the middle (6/10 and 8/10) demonstrating the highest scores on the three dependent variables.

The next goal was to analyze the differences in responses to the AC problems compared to the CA problems, and to also examine the degree to which the experimental manipulation facilitated solutions on those two types of problems. For all of these analyses, the dependent variable for complete falsification scores (ranging from 0-9) was used, since there was no interest in any response pattern except for the p, not-q selection. In order to avoid comparisons between six AC problems and three CA problems a composite score for the three CA problems was created and compared the CA composite score to composite scores representing every combination of three of the six AC problems, creating 20 triplets for use in the analysis. In table 3 the mean and standard deviation for the 20 AC triplets and 3 CA problems are displayed along with the results of the pairwise analyses.

Table 3. Composite Score Comparisons for AC and CA Problems

| Average | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|
| CA | 0.96 | 1.185 | 0.094 |
| AC1 | 2.45 | 0.869 | 0.069 |
| AC2 | 2.35 | 0.842 | 0.067 |
| AC3 | 2.36 | 0.838 | 0.066 |
| AC4 | 2.33 | 0.875 | 0.069 |
| AC5 | 2.49 | 0.856 | 0.068 |
| AC6 | 2.51 | 0.818 | 0.065 |
| AC7 | 2.47 | 0.848 | 0.067 |
| AC8 | 2.41 | 0.858 | 0.068 |
| AC9 | 2.37 | 0.925 | 0.073 |
| AC10 | 2.39 | 0.885 | 0.07 |
| AC11 | 2.3 | 0.919 | 0.073 |
| AC12 | 2.32 | 0.937 | 0.074 |
| AC13 | 2.28 | 0.942 | 0.075 |
| AC14 | 2.22 | 0.952 | 0.076 |
| AC15 | 2.18 | 0.993 | 0.079 |
| AC16 | 2.2 | 1.005 | 0.08 |
| AC17 | 2.36 | 0.931 | 0.074 |
| AC18 | 2.33 | 0.965 | 0.076 |
| AC19 | 2.35 | 0.948 | 0.075 |
| AC20 | 2.25 | 1.054 | 0.084 |

N=159, All comparisons between each AC triplet and the CA triplet were significant at $p < .000$

Every comparison between AC problems and CA problems was significant at p<.000. In order to evaluate the impact of the experimental manipulation on falsification scores, one-way ANOVAs were calculated for the CA composite score and a composite score of the 6 AC problems. The ANOVA for the CA problems was not significant: F (4,154) = 1.353, p=.253. The ANOVA for the AC problems was significant at F (4,154) = 5.180, *p=.001*. Post-HOC analyses revealed that scores for the control group (M 3.73, SD 1.97) were significantly lower than scores for the 6/10 group (M 5.13, SD 1.36, p=.037), and the 8/10 group (M 5.32, SD 1.08, p = .050). None of the criterion groups was significantly different than the others. The same inverted-U pattern was seen for AC problems as is presented above for the overall problem set. The pattern was not replicated for the CA problems.

Next, we calculated the correct solution means for each of the four argument forms (Modus Ponens-selecting the *P* card, Denial of Antecedent-not selecting the *not-P* card, Affirmation of Consequent-not selecting the Q *card*, Modus Tollens-selecting the *not-Q* card) were calculated. To evaluate the performance on argument forms, a pairwise samples test was computed, between the selection of each of the four argument forms on the three CA problems and each of the 20 AC triplets, using the number of correct answers to each argument form as the dependent measure. None of the 20 comparisons for Modus Ponens were significant, indicating that the Modus Ponens argument form is not affected by problem type. Of the 20 comparisons for Denial of Antecedent, 11 were significant, with the CA problems being more difficult in every case. Of the 20 comparisons for Affirmation of Consequent, 100% were significant at p<.000. The average score for CA problems was 1.23 out of 3 while the average scores for AC problems was 2.66. Similarly, of the 20 comparisons for Modus Tollens, 100% were significant at p<.000. The average score for CA problems was 1.53 while the average score for AC problems was 2.71.

In order to test the effect of the experimental manipulation on the correct selection of each of the four argument forms, one-way ANOVAs were calculated for the four CA composite scores (one for each argument form) and the four composite scores of the 6 AC problems. The following significant comparisons were found: 1) The ANOVA for the affirmation of consequent on the AC composite score: F (4,154) =4.160, p=.003. Posthoc comparisons showed that participants in the control condition (M 4.88, SD 1.45) scored significantly lower than participants in the 8/10 condition (M 5.71, SD .69, p=.027).

2) The ANOVA for Modus Tollens on the AC composite score: F (4,154) =3.767, p=.006. Posthoc comparisons showed that participants in the control condition (M 4.94, SD 1.37) scored significantly lower than participants in the 8/10 condition (M 5.87, SD .43, p=.005).

3) The ANOVA for the denial of antecedent on the CA composite score: F (4,154) =4.222, p=.003. Posthoc comparisons showed that participants in the 6/10 condition (M 2.22, SD 1.13) scored significantly lower than participants in the 4/10 condition (M 2.88, SD .49, p = .006), the 8/10 condition (M 2.77, SD .62, p=.036) and the 10/10 condition (M 2.81, SD .48, p = .021).

4.) The ANOVA for the Modus Tollens on the CA composite score: F (4,154) =3.405, p=.011. Posthoc comparisons showed that participants in the control condition (M 1.03, SD 1.10) scored significantly lower than participants in the 6/10 condition (M 2.06, SD 1.16, p=.007).

The patterns of card selection for each problem were also examined. There were 16 possible selection combinations for each problem. Although each of the 16 possible combinations occurred at least once, reasonably large frequencies were found only for the logically correct P, not-Q (67.08%); the *P* and *Q* (15.63%); the *P* (3.15%); and the *P*, *Q*, and *not*-*Q* (3.73%) selection combinations. The proportions of selection frequencies of these 4 combinations account for 89.59% of the total responding. The *P* and *Q* selection combination was chosen more often for the CA problems (39.30%), compared to the AC problems (3.79%)

In order to avoid comparisons between 6 AC problems and 3 CA problems, a composite score for the selection of the *P*, *Q* strategy on the 3 CA problems was created and compared the CA composite score to composite scores representing the selection of the *P*, *Q* strategy for every combination of 3 of the six AC problems, creating 20 triplets for use in the analysis. Every comparison between AC problems and CA problems was significant at p<.000. Out of a possible 3 points, the average number of selections of the P, Q strategy for AC problems was 0.0142, while for CA problems the average was .843. In order to evaluate the impact of the experimental manipulation on the selection of the *P*, *Q* strategy, one-way ANOVAs were calculated for the CA composite score and a composite score of the 6 AC problems. The ANOVA for the CA problems was not significant (*p=.449).* The ANOVA for the AC problems was also not significant (*p=.106).*

## Conclusion

## Discussion

One goal of this study was to investigate the impact of realistic expectations on an academic task. The results confirmed the main hypothesis, that knowledge of progressively increasing success criteria would produce an inverted-U shaped curve for correct problem solutions. In short, these results indicate that students can be motivated to attempt academic work when they know what is expected of them, especially when those expectations are realistic. It was attempted here to help evaluate whether or not knowledge of criteria can remove cognitive or motivational obstacles to the display of deductive reasoning competence.

As discussed above, it has been theorized that all adults possess the ability to use deductive reasoning. This study attempted to experimentally investigate practices professors can use to help students achieve their full potential. Educational intervention can take a long time to investigate. This task may be a way to examine the effectiveness of educational practices in a much shorter time frame. If knowledge of realistic criteria can improve performance on this task, future research on the use of criteria and expectations to elicit improved understanding and effort on the part of students can be suggested.

In general, this study fits in with a host of experiments underway in this Psychology department, which seek to provide experimental evidence of best teaching and assessment practices, as they pertain to students at one Historically Black College/University. An exciting prospect that this study suggests for the future of educational experimentation is that the selection task can be used to pilot interventions in a short period of time, in order to discover the potential of the researchers' hypotheses, before they are implemented on a larger scale.

The second goal of this study was to offer further support of the findings from Muller, Overton and Reene's (2001) study of AC and CA problems, and the differential success rates they elicit. Inspection of card selections across problems demonstrates that AC problems were more likely to lead to a correct search for counterexamples (*not*-*Q*). On AC problems, participants were also less likely to select the irrelevant *Q* card than on CA problems. On CA problems, a common error consisted of selecting the *Q* card, and in failing to select the *not*-*Q* card. The ability to produce alternate antecedents that are meaningfully related to the consequent has been identified as a prerequisite for the correct solution to the Affirmation of the Consequent invalid argument form (Markovits, 1993). In the selection task, this solution entails not choosing the *Q* card. Better performance on the AC problems than on the CA problems for the *Q* card may have been a result of the semantic format of the consequent terms involved in the rules and presented as cards. The specific character of the consequent terms may involve different levels of cognitive complexity and may pose different cognitive processing demands. Although the construction of alternative examples of *P* and *Q* for concrete problems may require only class inclusion reasoning (Kuhn, 1977; Piaget, 1974/1980), the construction of alternate examples of *P* and *Q* models for problems with a binary consequent (Barouillet and Lecas, 1998) or for more abstract problems may require the construction of a broader scope of negation (Byrnes, 1988; Piaget and

Garcia, 1991), increased flexibility in dividing and integrating logical classes, and the ability to generate multiple possibilities (Markovits, 1993).

A second explanation is that the relative ease of not choosing the *Q* card in AC as compared to CA problems may be due to the semantic relation between antecedent and consequent. A major problem in solving conditional reasoning problems consists of the ability to realize that an asymmetric relation holds between antecedent and consequent (i.e., if *P* then *Q*, but if *Q* then not necessarily *P*; see Staudenmayer and Bourne, 1977). Now, in AC problems, the consequent-clause is a precondition for the action or event mentioned in the antecedent. Thus, an AC problem states a relation of conditionship and can be translated as "*Q* is the necessary condition for *P*" (see Matalon, 1990) or "*P* only if *Q*" (Thompson, 1995, Experiment 1). The explicit statement of the precondition relation highlights the asymmetric nature of the relation between antecedent and consequent: Satisfying *Q* is only a necessary, but not sufficient condition, for the occurrence of *P*. By contrast, in the CA schema, the particular temporal relation between antecedent and consequent does not suggest an asymmetric relation between antecedent and consequent; adults are also unlikely to rephrase an obligation as "*P* only if *Q*" (Thompson, 1995, Experiment 1). These explanations for correctly not selecting the *Q* card are not necessarily mutually exclusive. Rather, the semantic format of the consequent term may facilitate the construction of an asymmetrical relation between antecedent and consequent. The relative contributions of the semantic format of the consequent term and of the semantic relation between antecedent and consequent to the suppression of the *Q* card could be examined by systematically controlling for the semantic format in AC and CA problems.

Now turn to the question of why the selection of the *not-Q* card is more difficult in AC than in CA problems. Note that the Modus Tollens inference is more difficult in the selection task than in the evaluation task. Whereas in the evaluation task participants are explicitly asked about the state of affairs that holds when *Q* does not occur, in the selection task participants must spontaneously discover the relevance of the *not-Q* card (Thompson, 1995). It is suggested that the relative ease of correctly selecting the *not-Q* card in AC as compared to CA problems is due to the semantic structure between antecedent and consequent. The semantic structure between antecedent and consequent in AC problems is such that it suggests that a necessary condition is not fulfilled (Matalon, 1990). As a result, participants do not have to construct the relevance of the *not-Q* card inferentially; rather, its relevance is established by the semantic relation between antecedent and consequent. This contrasts with CA problems in which the temporal relation between antecedent and consequent makes it more difficult to realize that the occurrence of the consequent is a necessary condition for the occurrence of the antecedent. As a consequence, the construction of the relevance of the *not-Q* card may require more complex logical operations. In particular, it may require the comprehension of the law of duality, which states that the complement class of a superordinate class is smaller than the complement class of a subordinate class (Inhelder and Piaget, 1958). For example, the class of things that are not birds is smaller than the class of things that are not eagles. It is likely that understanding the law of duality is required for choosing the *not-Q* card because participants must recognize that if the antecedent (*P*) is included in the consequent (*Q*) then the complement of the including class (*not-Q*) cannot be included in the complement of the included class (*not-P*).

Recommendations for future research include the continued use of the selection task to investigate educational interventions. There is also are interest in studies which would apply item analyses to the ten problems, in order to potentially rework the measure to include items which can provide more reliable and valid results vis-à-vis logical competence and the real-time obstacles to exhibiting that competence.

**References:**

Allen, D., and Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit forboth teachers and learners. *Life Sciences Education, 5*, 197-203.

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership,* 57 (5), 13-18.

Andrade, H. and Saddler, B. (2004). The writing rubric. *Educational Leadership*, 62 (2),48-52.

Arter, J., and McTighe, J. (2001). *Scoring rubrics in the classroom*. Thousand Oaks, CA: Corwin Press.

Babad, E. (1998). Preferential affect: the crux of the teacher expectancy issue. In J. Brophy (ed.), *Advances in Research on Teaching: Expectations in the Classroom* VII, Greenwich CT: JAI Press, pp. 183-214.

Barouillet, P.,and Lecas, J. -F. (1998). How can mental models theory account for content effects in conditional reasoning? A developmental perspective. *Cognition, 67,* 209–253.

Bissell, A. N., and Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *Bioscience, 56*, 66-72.

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*, 631-61.

Byrnes, J. P., and Overton, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology, 22(6*), 793–799.

Chappell, M., and Overton, W. F. (1998). Development of logical reasoning in the context of parental style and test anxiety. *Merrill-Palmer Quarterly, 44,* 141–156.
Chappell, M.S. and Overton, W.F. (2002). Development of logical reasoning and the school performance of African-American adolescents in relation to socioeconomic status, ethnic identity, and self-esteem. *Journal of Black Psychology, 28*, 295-317.
Cheng PW and Holyoak KJ. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391-416.
Cooper, H. M. (1985). Models of teacher expectation communication. In J. B. Dusek (ed.), *Teacher Expectancies,* Hillsdale NJ: Erlbaum, pp. 135-58.
Dahm, K. D., Newell, J. A., and Newell, H. L. (2004). Rubric development for assessment of undergraduate research: Evaluating multidisciplinary team projects. *Chemical Engineering Education, 38* (1), 68-73.
Dochy, F., Gijbels, D., and Segers, M. (2006). Learning and the emerging new assessment culture.new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts, and S. Vosniadou Eds.), *Instructional psychology: Past, present, and future trends* (pp. 191-206). Oxford, Amsterdam: Elsevier.
Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology. 31*, 674-685.
Eddy , S. (2010, November 03). *Psycho hawks*. Retrieved from http://psychohawks.wordpress.com/2010/11/03/theories-of-cognitive-development-lev-vygotsky/
Feldman, R. S., and Theiss, A. J. (1982). The teacher and student as Pygmalions: Joint effects of teacher and student expectations. *Journal of Educational Psychology, 74,* 217-23.
Frederiksen, J. R., and Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*, 27-32.
Good, T. L., and Weinstein, R. S. (1986). Classroom expectations: one framework for exploring expectations. In K. K. Zumwalt (ed.), *1986* ASCD *Yearbook: Improving Teaching,* Alexandria VA: Association for Supervision and Curriculum Development.
Goodrich, H. (1997). Understanding rubrics. *Educational Leadership,* 54 (4), 14-17.
Grant, H. and Dweck, C.S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*, 541-553.
Haynes, J., Tikly, L., and Caballero, C. (2006). The barriers to achievement for white/black Caribbean pupils in English schools. *British Journal of Sociology of Education,* 27 (5), 569-83.
Hong, Y., Chiu, C., Dweck, C.S., Lin, D.M. and Wan, W. (1999). Implicit Theories, Attributions, and Coping: Meaning System Approach. *Journal of*

*Personality and Social Psychology*, *77*, 588-599.
Howell, R. J. (2011). Exploring the impact of grading rubrics on academic performance: Findings from a quasi-experimental, pre-post evaluation. *Journal on Excellence in College Teaching*, *22* (2), 31-49.
Huba, Mary and Jann Freed (2000). *Learner-Centered Assessment on College Campuses*. Allyn and Bacon: Boston.
Inhelder, B. and Piaget, J. (1958). *The growth of logical thinking: from childhood to adolescence.* New York: Basic Books.
Johnson-Laird PN. (2006). *How We Reason.* New York: Oxford University Press.
Kuhn D, Franklin S. (2006). The second decade: What develops (and how). In: Kuhn D, Siegler RS, eds. *Cognition, perception, and language: Vol. 2. Handbook of child psychology (6th ed.).* Editors-in –chief Damon W, Lerner RM. Hoboken, NJ: Wiley.
Markovits, H. (1993). The development of conditional reasoning: A Piagetian reformulation of mental models theory. *Merrill-Palmer Quarterly, 39,* 131–158.
Markovits H, Barrouillet P. (2002). The development of conditional reasoning: A mental model account. *Developmental Review, 22,* 5-36.
Matalon, B. (1990). A genetic study of implication. In W. F. Overton (Ed.), *Reasoning, necessity, and logic* (pp. 87–110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. (Original work published 1962)
McWhirter, E. H., Torres, D. M., Salgado, S., and Valdez, M. (2007). Perceived barriers and postsecondary plans in Mexican American and white adolescents. *Journal of Career Assessment 15 (*1) 119-38.
Moshman D. (2004). From inference to reasoning: The construction of rationality. *Thinking and Reasoning*, 10, 221-241.
Moskal, B.M. (2000). Scoring Rubrics: What, When and How? *Practical Assessment,Research and Evaluation*, *1* (3). Retrieved July 13, 2008 from http://pareonline.net/getvn.asp?v=7andn=10 measure student outcomes. *Journal of Public Affairs Education*, *12* (3), 295-311.
Muller, U., Overton, W.F., and Reene, K. (2001). Development of conditional reasoning: A longitudinal study. *Journal of Cognition and Development*, *2*, 27-49.
O'Brien D, Overton WF. (1982). Conditional reasoning and the competence-performance issue: A developmental analysis of a training task. *Journal of Experimental Child Psychology, 34*, 274-290.
Overton, W.F. (1990). Competence and procedures: Constraints on the development of logical reasoning.In W.F. Overton (Ed.), *Reasoning, Necessity and Logic: Developmental Perspectives*. Hillsdale, NJ Erlbaum.

Overton, W. F. (Ed). (2010) Cognition, biology, and methods across the lifespan. Volume 1 of the Handbook of life-span development. Editor-in-chief: R. M. Lerner. Hoboken, NJ: Wiley.

Overton, W. F., Byrnes, J. P., and O'Brien, D. P. (1985). Developmental and individual differences in conditional reasoning: The role of contradiction training and cognitive style. *Developmental Psychology, 21*(4), 692–701.

Overton WF, Newman J. Cognitive development: A competence-activation/utilization approach. In: Field T, Huston A, Quay H, Troll L, Finley G, eds. *Review of Human Development*. New York: Wiley and Sons, Inc; 1982, 217-241.

Overton, W. F., Ward, S. L., Noveck, I. A., Black, J., and O'Brien, D. P. (1987). Form and content in the development of deductive reasoning. *Developmental Psychology, 23*, 22–30.

Perlman, C.L. (2003). Performance assessment: Designing appropriate performance tasks and scoring rubrics. In C. Boston (Ed.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 497-506). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Plaks, J.E., Grant, H. and Dweck, C. (2005). Violations of implicit theories and the sense of prediction and control: Implications for motivated person perception. *Journal of Personality and Social Psychology, 88,* 245-262.

Ricco, R.B. (2010). Development of deductive reasoning across the life span. In W.F. Overton (Ed.), *Cognition, biology, and methods across the lifespan*. Vol. q of the Handbook of life-span development. Editor-in-chief: R.M. Lerner. Hoboken, NJ: Wiley.

Ricco, R. and Overton, W. F. (2011). Dual systems competence ←-→ procedural processing: A relational developmental systems approach to reasoning. *Developmental Review,* 31, 119-150.

Rich, J.D., Fullard, W. and Overton, W. (2011). The relationship between deductive reasoning ability, test anxiety and standardized test scores in a Latino sample. *Hispanic Journal of Behavioral Sciences, 33* 261-277.

Rubie-Davies, C. M. (2006). 'Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools, 43 (5),* 537-52.

Rubie-Davies, C.M. (2007). Classroom interactions: Exploring the practices of high and low-expectation teachers. *British Journal of Educational Psychology, 77* (2), 289-306.

Rubie-Davies, C.M., Peterson, E., Irving, E., Widdowson, D. and Dixon, R. (2010). Expectations of achievement: Student, teacher and parent perceptions. *Research in Education, 83,* 36-53.

Staudenmayer, H., and Bourne, L. E., Jr. (1977). Learning to interpret conditional sentences: A developmental study. *Developmental Psychology, 13,* 616–623.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49,* 1–58.

Wason, P.C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135-151). Harmondsworth: Penguin.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. S. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge and Kegan Paul.

Weinstein, R. S. (2002), *Reaching Higher: The Power of Expectations in Schooling,* Cambridge MA: Harvard University Press.

Wiener, B. (1980). The role of affect in rational (attributional) approaches to human motivation. *Educational Research, 9,* 4-11.

Wiener, B. (1982). The emotional consequences of causal ascriptions. In M.S. Clark and S.T. Fiske (Eds.), *Affect and cognition: The 17th annual Carnegie symposium on cognition* (pp. 185-208). Hillsdale, NJ: Erlbaum.