Reflective Judgment: Concepts of Justification and Their Relationship to Age and Education

Karen Strohm Kitchener
School of Education
University of Denver

Patricia M. King
Office of Student Services
Ohio State University

This paper describes a seven-stage model of post-adolescent reasoning styles, the most advanced of which is called "Reflective Judgment." The model outlines a sequence of increasingly complex methods of justification of beliefs. This model demonstrates how people's conceptions of the nature of knowledge, the nature of reality, and their concepts of justification change over age/educational levels. Sixty subjects from three age/educational levels (high school, college, and graduate school) were administered the Reflective Judgment Interview. In addition, subjects were tested on four other factors hypothesized to affect reflective judgment scores. Two were competing theoretical constructs: verbal ability (using Terman's Concept Mastery Test) and Piagetian formal operations (chemicals and pendulum tasks); and two were potentially confounding factors: socio-economic status (using Hollingshead's two-factor index) and verbal fluency (number of words spoken during the interview). Highly significant differences (p < .001) were found on reflective judgment level between the three age/educational groups, and could not be statistically accounted for by scores earned on measures of the other four factors. Whereas verbal ability was found to be closely related to reflective judgment level, the differences between groups on reflective judgment could not be solely attributed to this factor.

People are often faced with the task of making judgments about controversial issues, such as assessing the danger of nuclear energy, deciding whether to purchase beverages containing saccharin, or whether to believe a news story.

*The authors wish to thank Mark L. Davison and Clyde A. Parker for their contributions to conceptualization and design as well as for careful reading of the manuscript. They wish also to extend their thanks to T. Bouchard, A. Collins, D. Keating, and J. Rest for reading the original manuscripts, and to R. Kitchener and M. Brabeck for their many suggestions in developing and refining the model. This study was conducted at the University of Minnesota and was partially supported by the University of Minnesota Computer Center. Authors contributed equally to collection of data and writing of manuscript.
Discussing one’s point of view with others is a familiar part of daily conversation: “What did you think about __________?”. Learning how to convincingly defend one’s point of view to others is central to the educational mission of many secondary and post-secondary educational institutions. How, then, do people explain and justify their beliefs to others? How does one come to hold a particular point of view about an issue, and why does an individual hold that view instead of a different one? Consider the following statements given to defend a point of view:

**Interviewer:** What do you think about this issue (special creation versus evolution)?

**Subject:** I believe in my religion. God created man as a person. We did not evolve.

**I:** On what do you base that point of view?

**S:** On my religious background, what I’ve been told.

**I:** Can you tell me more about how that happened, the process of learning about that?

**S:** I was taught by my parents and the church. It’s worked into me that way.

—Subject A

**Interviewer:** When people differ about matters such as this, how does one make a determination about what is correct?

**Subject:** Again, how well thought-out the positions are, at what level one chooses to argue and support the position, what kinds of reasoning and evidence one would use to support it, how it fits into the rest of one’s world view or rational explanations . . . , how consistent the way in which one argues on this issue is as compared with other issues.

—Subject B

These examples show widely divergent means by which people justify their points of view, from unexamined reliance on the word of an authority figure to a thoughtful examination and evaluation of the available evidence and a determination of a reasoned judgment based on that evidence. In this paper it will be argued that differences in concepts of justification, such as those illustrated above, are derived from different assumptions about reality and knowledge. In addition, a study will be reported that investigated differences in concepts of justification across age and educational levels in a sample of young adults.

Little has been written about development in reasoning beyond adolescence, perhaps because of the Inhelder and Piaget (1958) claim that formal operations was the final form of equilibrium in mental operations. Recently, others (Arlin, 1975; Broughton, 1975, 1978; Gilligan & Murphy, 1979;
Kohlberg, 1973; Kuhn, 1979; Reigel, 1973) have questioned whether the real life reasoning of adults is adequately encompassed by the hypothetico-deductive reasoning of formal operations, and have argued for alternative models. Kitchener and Kitchener (1981) have argued that while different people use hypothetico-deductive reasoning in equally correct ways, they still draw widely varying conclusions about controversial issues. Further, they have shown that these differences in their conclusions are related to their prior assumptions and criteria for evaluating arguments, and not to formal logic.

Perry (1968, 1970) describes a model of intellectual and ethical development in early adulthood which is based on his observations of college students. Using a nine-position model, he traces development from egocentric absolutism to contextual relativism to commitment in relativism. Positions 1 through 5 are characterized by increasing complexity in decision making and view of knowledge, and a decreasing dependence on authority figures in decision making. The highest four positions, 6 through 9, are characterized by the focus on individual responsibility through the making of commitments in relativism. Such commitments are made after the active consideration of alternative choices. Perry characterizes these commitments as occurring in such areas as vocation and marriage. Evidence has been presented that college seniors score in higher Perry positions than do freshmen (Meyer, 1977; Perry, 1968) and that development can be systematically promoted by educational interventions (Clinchy, Lief, & Young, 1977; Stephenson & Hunt, 1977; Touchton, Wertheimer, Cornfeld, & Harrison, 1977; Widick, Knefelkamp, & Parker, 1975). The work of Kurfiss (1975, 1977) supports the sequentiality of positions 2 through 6, but not their structural unity. The evidence based on Perry’s model suggests that assumptions about knowledge are related to forms of decision making and that these develop during the college years from positions 2 through 5. However, because of the apparent shift in focus between positions 5 and 6, from intellectual to identity development, it is very difficult to test whether further development occurs in assumptions about knowledge and in forms of decision making beyond position 5.

THE DEVELOPMENT OF REFLECTIVE JUDGMENT

Building on Perry’s model, but also drawing from the work of Harvey, Hunt, and Schroder (1961), Loevinger (1976), Reigel (1973), and Broughton (1975), a seven-stage model explicating the development of concepts of justification was created (King, 1977; Kitchener, 1978). This model outlines a sequence of increasingly complex assumptions about knowledge and reality and shows how they are related to forms of justification.

Each stage represents a logically coherent network of assumptions and corresponding concepts that are used to justify beliefs. For example, if one
believes that authority figures are the source of absolute knowledge, then it is reasonable to justify one's beliefs by simple reference to an authority's beliefs ("what is true for an authority figure is true for me"). By contrast, if one believes that absolute knowledge is for practical reasons impossible to obtain and that knowledge is idiosyncratic to the individual, then it follows that beliefs will be justified based on an idiosyncratic evaluation of data ("what is true for me is not necessarily true for others"); that is, it would be inconsistent and illogical to believe both that authority figures are the source of absolute knowledge and that knowledge is idiosyncratic to the individual. In other words, the different assumptions about reality and knowledge at each stage imply different forms of justification. At each stage the network of assumptions provides categories which the individual uses to perceive and organize available information and to make judgments about an issue. The process of forming judgments becomes increasingly more complex, sophisticated and comprehensive from lower to higher stages. For example, in the lower stages (1 and 2), authorities serve as sources of truth, and therefore correspondence with authoritative views is the major criterion used for justifying one's own beliefs. At the highest stage (7), by contrast, reality is understood as existing objectively, but the process of inquiry is seen as being fallible. The major criterion for justification at this stage is the degree to which the belief is based on a rational process of conjecture that demonstrates the use of evidence and rules of inquiry appropriate for the issue at hand. This form of justification is based on the principle that knowledge statements must be evaluated as more or less likely approximations to the truth and that they must be open to the scrutiny and criticisms of other rational people; this form is termed "reflective judgment." "Reflective judgment model" refers to the seven-stage sequence leading to the ability to make reflective judgments. "Reflective judgment stage" refers to the degree to which a person is able to make reflective judgments.

In the following section the major assumptions about reality (R) and the assumptions about knowledge (K) of each stage will be identified, along with the corresponding concepts of justification. An example will be given which illustrates a major characteristic of the stage. These examples are verbatim statements which were made in response to intellectual controversies presented to the subjects. The controversies covered the following topics: special creation vs. evolution, the building of the pyramids, the objectivity of news reporting and the danger of chemical additives in foods. (See the methods section for a listing of one such controversy.)

Stage 1

Assumption R. There is an objective reality which exists as the individual sees it. Reality and knowledge about reality are identical and known absolutely through the individual's perceptions.
Assumption K. Knowledge exists absolutely. One's own views and those of authorities are assumed to correspond to each other and to absolute knowledge. Knowledge is gained through the individual's perceptions and prior teaching.

Concept of Justification. Beliefs simply exist; they are not derived and need not be explained. Differences in opinion are not perceived, and justification is therefore unnecessary.

Example 1.

S: Somebody had to start this world, so I have to believe in my religion that there is a God.
I: Have you ever doubted that?
S: No, I am not a religious freak or nothing. I just go to church and just think it's common knowledge. Somebody had to start it.

In this example, doubt is precluded with the pre-existing belief that "somebody had to start it." Alternatives do not exist for this individual since the "truth" is common knowledge. There is a failure in Stage 1 to differentiate one's own beliefs from the beliefs of an authority. In this example, the authority figure is defined as the church, not taking into consideration that other churches may espouse different doctrines. Those people who may be objectively defined as authorities in their field or who are in other positions of power or influence may initially be discredited or dismissed if their views do not coincide with the individual's own views. When personal authorities do not have answers or when there is disagreement among authorities about the "true" answers for an issue, the individual must then deny the discrepancy or re-evaluate the assumption that absolute knowledge and one's own view of reality are identical. Seeking answers elsewhere, even from other authorities, forces movement beyond the simple egocentricism of Stage 1.

Stage 2

Assumption R. There is an objective reality which is knowable and known by someone.

Assumption K. Absolute knowledge exists, but it may not be immediately available to the individual. It is, however, available to legitimate authorities.

Concept of Justification. Beliefs either exist or are based on the absolute knowledge of a legitimate authority.
Example 2.

S: How can two scientists go through the same procedures and come up with different results?
I: How can they?
S: It's like the commercials, like big companies such as Bayer or Excedrin... I just saw a commercial on it. Their scientists study this and they are being paid to say it's okay!
I: What if neutral experts disagree?
S: If they both did it the same way, if they use the same chemicals and everything, I don't think it could have come out differently. If they did, then somebody probably did something wrong, like add something extra.

In this example, the individual acknowledges that scientists (the "authorities" on this issue) may disagree, but implies that is because they are dishonest or that they "fix" the data. If honest scientists went about the procedure in the correct way, the person implies, there would be no disagreement between conclusions.

Allowing for the possibility that problems exist for which the individual has no answers forces a person to abandon the view that the world is only as she/he perceives it. The individual must as a result seek answers to questions and not rely exclusively on his or her own prior perceptions and beliefs. However, because individuals who reason in this manner maintain the belief that knowledge is both certain and possible to attain, they also assume that someone must have the knowledge, even if they personally are ignorant or poorly informed. The source of certain knowledge remains legitimate authorities. In the process of identifying authorities who have absolute knowledge, the individual discovers that there are people in positions of authority who disagree with each other and with the individual's preconceived understanding of the world. Authorities, such as the scientists described in Example 2, are then divided into two groups: legitimate and illegitimate authorities. This change reflects a development from belief in singular pre-existing knowledge to a belief in two categories of knowledge—that which is right and that which is wrong. These changes are an advancement from Stage 1 since they allow alternative conceptions of reality to be accommodated as wrong or misguided.

With the realization that there is an alternative perspective from which to understand a problem lies the possibility that there may be even more than one such alternative. Second, in the realization that there are right authorities and wrong others lies the possibility that even legitimate authorities may disagree among themselves. Third, in debates over the extent to which one's own view is right and another is wrong lies the possibility that the other's view may appear reasonable and appealing, and that one's own view may have to be explained and justified. These possibilities, as well as the real diversity and discrepancy in actual experiences that occur in a pluralistic environment, cannot be accommo-
dated by the assumptions of Stage 2. The need for a new solution emerges and leads to the abandonment of the view that certain knowledge is always possible to attain.

Stage 3

Assumption R. There is an objective reality, but it cannot always be immediately known, even to legitimate authorities. It is possible to attain knowledge about this reality, but our full knowledge of it is as yet incomplete and therefore uncertain.

Assumption K. Absolute knowledge exists in some areas, but in others it is uncertain, at least temporarily. Even authorities may not have certain knowledge, and therefore cannot always be depended upon as sources of knowledge. Knowledge is manifest in evidence which is understood in a concrete, quantitative way such that a large accumulation of evidence will lead to absolute truth.

Concept of Justification. Beliefs either exist or are based on an accumulation of evidence that leads to absolute knowledge. When such evidence is not available, individuals claim that while waiting for absolute knowledge to become available, people can temporarily believe whatever they choose to believe.

Example 3.

I: Is one point of view right and one wrong?  
S: I could not say one is right. That is why I said for you to get your information and I will get mine and we will compare it and see what comes out. I am a middle-of-the-road person. I can be changed one way or the other, whatever comes along. I can see if it's true or looks to be that way. I am open minded enough to listen.

In Example 3, the individual acknowledges that there is no singular right belief but does not offer a rational explanation for his beliefs. At this stage, there is an assumption that there is no adequate way to judge or to evaluate a point of view, except for the quantity of information available.

In the progression from Stages 1 to 3, there is an increase in the number of categories available for processing different perceptions of the world, from one to two to many. This change is reflected in the assumption of Stage 3 that "there are many ways to perceive the world and mine is only one." This assumption, along with the view that authorities are fallible, leaves the individual vulnerable when trying to justify his or her beliefs. Although quantity of evidence is seen as the last stronghold of absolute knowledge ("with enough evidence we will know for sure"), its unavailability leaves the individual with no absolute basis for
forming decisions about pressing, current issues. She or he thus concludes that until the truth emerges, it is as accurate to believe one thing as another, and therefore it does not matter what you believe.

In the Stage 3 admission that in some cases knowledge is temporarily uncertain lies the potential for recognizing that uncertainty is an important part of the process of inquiry, a characteristic of Stage 4. In addition, the dissonance created by holding unjustified beliefs in light of being asked for justification (such as in school settings) leaves one to seek a more defensible principle on which to base one’s belief.

Stage 4

_Assumption R._ There is an objective reality, but it can never be known without uncertainty. Neither authorities, time or money nor a quantity of evidence can be relied upon to ultimately lead to absolute knowledge.

_Assumption K._ Absolute knowledge is for practical reasons impossible to attain, and is therefore always uncertain. There are many possible answers to every question, but without certainty and a way to adjudicate between answers, there is no way to decide which one is correct, or even whether one is better than another. Knowledge is idiosyncratic to the individual.

_Concept of Justification._ Beliefs are justified with idiosyncratic knowledge claims and on idiosyncratic evaluations of data ("what is true is true for me, but not necessarily for everyone else"). The individual is the ultimate source and judge of his or her own truth.

_Example 4._

_I:_ Can you say one point of view is better and another worse?
_S:_ No, I really can’t on this issue (creation vs. evolution). It depends on your belief, since there is no way of proving either one.
_I:_ Can you say that one is more accurate than the other?
_S:_ No, I can’t. I believe they are both the same as far as accuracy.
_I:_ Would you go so far as to say your opinion is the right one?
_S:_ No.
_I:_ But yet you believe so strongly in it—that’s why I am asking . . .
_S:_ I am the type of person who would never tell anyone that their idea is wrong—if they had searched, well, even if they hadn’t searched, even if they just believe it—that’s cool for them.
_I:_ Can you say that one opinion is better and one is worse?
_S:_ No, not at all. It’s better for them and like their opinion would probably be worse for me.
Individuals in Stage 4 become practical skeptics. They assume that uncertainty will always exist because for practical reasons (such as lack of complete and accurate historical records or inaccessibility of some types of scientific data) we can never have perfect knowledge of objective reality. Waiting for the right answer to emerge, as individuals do in Stage 3, does not suffice. Knowledge is limited to individualized claims about reality based on idiosyncratic assessments of evidence. Nothing can be judged or evaluated outside the individual's perspective because of the uncertainty of the relationship between knowledge and evidence. Here, all truth must be related to individual context and no evaluative judgments can be made between individual beliefs.

As people whose thinking is characteristic of Stage 4 debate about differences in views, and as they learn the rules of evidence for a particular domain (e.g., science, religion, history), they are eventually faced with the apparent reasonableness of many perspectives when evaluated in terms of evidence and other criteria of evaluation. This leads them to a better understanding of the relationship between knowledge and the rules of inquiry and from there, into Stage 5.

**Stage 5**

*Assumption R.* An objective understanding of reality is not possible since objective knowledge does not exist. Reality exists only subjectively, and what is known of reality reflects a strictly personal knowledge. Since objective reality does not exist, an objective understanding of reality is not possible.

*Assumption K.* Knowledge is subjective. Knowledge claims are limited to subjective interpretations from a particular perspective based on the rules of inquiry and of evaluation compatible with that perspective.

*Concept of Justification.* Beliefs are justified with appropriate decision rules for a particular perspective or context, e.g., that a simpler scientific theory is better than a complex one.

*Example 5.*

*S:* On that particular problem, I'd need to be more acquainted with what goes into building pyramids to take a point of view, because . . . when they say Egyptians did not have the mathematical knowledge or the tools, I wonder what evidence they are basing it on. Because it's quite possible they didn't find the evidence yet . . . so I'd tend to go along with the first point of view that the Egyptians built the pyramids because it is the simplest explanation at this point . . . because you have to come up with some far-out explanation of who built the pyramids if the Egyptians didn't. When you are uncertain about both, take the simplest explanation.
In the above example, the individual uses the principle of simplicity to decide between points of view, but does not offer an integrated point of view of his/her own. This is typical of the reasoning of Stage 5 since knowledge claims cannot extend beyond the rules of inquiry for the context (in this case, history and historiography). Rules of inquiry or evaluation do allow arguments to be evaluated as stronger or weaker, and this is a major advancement over Stage 4. When there is no principle of evaluation to cover a particular case, the individual equivocates and reflects upon the context of the particular situation or suggests that there is not enough information on which to decide.

The uncertainty of knowledge is accorded full status in Stage 5 not as a sign of the deficiency of knowledge, but rather as an essential aspect of it. Knowledge is subjective. The assumption of subjective knowledge presumes there are as many conceptions of reality as there are people perceiving it. Since knowledge is subjective, the individual seeks understanding by examining different conceptual frameworks or perspectives and by cautiously testing each perspective against empirical reality and the judgments of others. With the abandonment of absolute knowledge, individuals may engage in this examination with a new detachment which was not possible at previous stages.

The act of trying on one frame of reference and comparing it with another is at first based on simple rules of inquiry. However, it lays the groundwork for the formation of new and larger categories, categories which take into account the way frames of reference are similar and different. Harvey, Hunt, and Schroder (1961) have suggested that the evolution of complex, integrative thinking involves the development of “‘internal and informationally based concepts that act as a referent for evaluating feedback from the environment’” (p. 90). It is in the act of comparing perspectives against each other as well as against one’s personal experience that these new internal categories begin to emerge and form the basis of Stages 6 and 7.

Stage 6

**Assumption R.** An objective understanding of reality is not possible since our knowledge of reality is subject to our own perceptions and interpretations. However, some judgments about reality may be evaluated as more rational or based on stronger evidence than other judgments.

**Assumption K.** Objective knowledge is not possible to attain because our knowledge is based on subjective perceptions and interpretations. Knowledge claims can be constructed through generalized principles of inquiry and by abstracting common elements across different perspectives. The knower must play an active role in the construction of such claims.

**Concept of Justification.** Beliefs are justified for a particular issue by using generalized rules of evidence and inquiry. However, since our understanding of reality is subjective, any such justification is limited to a particular case, time or issue.
Example 6.

S: Well, I'm much more inclined to come down on the side that there's no such thing as unbiased, objective reporting. The very issue you view as being important, that you select out from other issues as being important, means that you've made something of a subjective choice. And from there, selecting even the minor facts that go in a story involves some kind of subjective choice. So that I don't think that there can be anything approaching objective reports of news events.

I: On what do you base your point of view?
S: It is based on the assumption that inevitably there has to be selection, and certainly there are overwhelming news events such as an earthquake or something of that nature which can be reported objectively, but nonetheless, what that represents is a subjective decision on the part of a lot of people that that's important.

Once individuals in Stage 5 give up the remnants of the view that knowledge has an absolute foundation, they eventually see that the spectator view of the knower must also be abandoned, and that the person must be involved in constructing a defensible view of reality. This occurs in Stage 6 in the acknowledgment that problematic situations require active thinking on the part of the individual. This increased responsibility for deciding what one believes beyond what different perspectives offer leads the individual to abstract common elements and principles of inquiry from different perspectives or contexts. These common elements allow an initial, tentative basis for objective evaluation over and above a particular perspective, using internalized categories of comparison and evaluation. Individuals in Stage 6 thus suggest that some knowledge claims may be warranted by the self-corrective process of inquiry itself, and, in particular, by the validity of the procedures that are used. However, since they believe that an objective understanding of reality is not possible, they will not argue that a particular solution offers the best rational conjecture about reality. For without a belief in objective reality, there is no standard against which viewpoints can be judged.

On the other hand, the tentative judgments, characteristic of Stage 6, initiate the process that will generalize in Stage 7 to an ability to synthesize elements into a coherent point of view, based on what Harvey, Hunt and Schroder called "abstract internal referents" (p. 91) of validity and probability.

Stage 7

Assumption R. There is an objective reality against which ideas and assumptions must ultimately be tested. Despite the fact that our knowledge of reality is subject to our own perceptions and interpretations, it is nevertheless possible, through the process of critical inquiry and evaluation, to determine that some judgments about that reality are more correct than other judgments.
**Assumption K.** Knowledge is the outcome of the on-going process of reasonable inquiry which ultimately leads to a concordence between belief and reality. The process of inquiry, however, may not always lead to correct claims about the nature of reality since the process itself is fallible. Knowledge statements must be evaluated as more or less likely approximations to reality and must be open to the scrutiny and criticisms of other rational people.

**Concept of Justification.** Beliefs reflect solutions that can be justified as most reasonable using general rules of inquiry or evaluation. Criteria for evaluation may vary from domain to domain (e.g., religion, literature, science), but the assumption that ideas, beliefs, etc., may be judged as better or worse approximations to reality remains constant.

**Example 7.**

*S:* It's my belief that you have to be very skeptical about what you read for popular consumption . . . even for professional consumption.

*I:* How do you ever know what to believe?

*S:* I read widely . . . of many points of view. Partly (it's) reliance on people you think you can rely on, who seem to be reputable journalists, who make measured judgments, then reading widely and estimating where the reputable people line up or where the weight of the evidence lies.

Same person reflecting about another issue:

*S:* It's (the view that the Egyptians built the pyramids) very far along the continuum of what is probable.

*I:* Can you say one (point of view) is right and one is wrong?

*S:* Right and wrong are not comfortable categories to assign to this kind of item . . . more or less likely or reasonable . . . more or less in keeping with what the facts seem to be.

In this example, and in Example B above, the individuals are able to evaluate their own beliefs as rational conjectures about reality, and they claim that their views are better approximations of reality than are other views. This process includes comparison against internalized categories of evaluation which have developed both through the formal process of education, such as learning how to judge hypotheses in science, religion, etc., and through experiencing the contradictions and compromises actually involved in working in these fields. These categories are those of “probably so,” “that which best fits reality,” “that which is discrepant from regularly observed processes,” etc. Beliefs are also evaluated against external criteria of validity and objectivity, such as what is admissible as evidence. These categories, along with the basic assumptions of Stage 7, make the individual readily able to fully engage in the process of
inquiry. Since individuals see their own ideas as subject to the criticisms of others, they may become involved in the critique of their own thinking. As they evaluate their own ideas, as well as the ideas of other people, issues which at one time appeared fully explained stand out from current categories as discrepant. These discrepancies provide the impetus for generating better and more complete explanations of reality. Individuals are thus actively engaged in the search for knowledge based on the assumption that an objective understanding of reality is possible to attain, and that they can evaluate and improve their own process of inquiry.

The reflective judgment model hypothesizes that there is sequential development in the forms and adequacy with which people justify their beliefs and that this development reflects concurrent development in the form and adequacy of their assumptions about reality and knowledge. In other words, higher stages of reflective judgment develop out of lower stages in a sequential fashion. Changes from stage to stage do not merely reflect different choices about what to believe about knowledge and reality nor new learning, but reflect changes in internalized categories of justification. What emerges in the higher stages (5–7) and is not present in the lower stages (1 and 2) is an ability to consciously reflect upon one’s own problem solving. This suggests that beliefs about particular problems are more a consequence of conscious choice in the higher stages than they are in the earlier ones. One result of the development of this quality is the ability to step back and evaluate one’s own solutions to problems. This supports the notion that the ability to monitor the outcome of one’s own cognitive processes when solving abstract intellectual problems (Flavell, 1979) is an emergent quality and is related to more basic changes in assumptions about reality and knowledge. It should be noted that there is no absolute dividing point in this scheme where this “reflective” ability is present or absent. Clearly, it is present in the higher stages and absent in the lower ones, and exists to a moderate degree in the middle positions.

Another component of these stages which should be noted is that each form of justification carries with it a set of logically related meta-cognitive beliefs, including such elements as what data are relevant to the problem-solving enterprise, how discrepant data ought to be viewed, and whether differences in opinion reflect crass bias or legitimate differences in interpretation of data. This suggests that there is a interactive relationship between more basic cognitive assumptions and meta-cognitive beliefs.

An initial test of this model was conducted using a series of pilot interviews with 18 subjects between 18 and 35 years old (reported in King, 1977, and Kitchener, 1978). These interviews consisted of a semi-structured discussion of six intellectual dilemmas, including scientific, religious and historical issues. The interviews were 45 to 90 minutes in length. In these discussions, each dilemma was presented in a standardized form, followed by a series of standardized probe questions. Pilot data confirmed that subjects’ responses could be
scored at each of the seven reflective judgment stages and that older subjects scored at higher stages than did younger ones.

The intent of the current study was to evaluate whether concepts of justification, as elaborated in the reflective judgment model were related to age and education. Two questions were investigated:

(a) Do reflective judgment stages differ across age/educational levels, with older, better educated subjects scoring at higher reflective judgment stages than younger subjects with less education?

(b) If so, can these differences be accounted for by other factors which are also related to intellectual development? Four such factors were investigated, two theoretical constructs (verbal ability and formal operations) and two potentially confounding factors (verbal fluency and socio-economic status).

Verbal ability has frequently been used as a measure of intellectual development. Previous longitudinal data suggest that verbal ability may increase or at least remain stable in the young adult years (Bayley, 1955; Bayley & Oden, 1955; Honzik & MacFarlane, 1973; Kansas & Bradway, 1971). Although there is some debate over interpretation of the increase in scores of measures of intellectual ability over time (Horn & Donaldson, 1976; Schaie, 1974), it was important to ask whether differences in reflective judgment could be attributable to age-related differences in verbal ability.

Formal reasoning has been described as "the crowning achievement of intellectual development" (Flavell, 1963, p. 202), but a sizeable proportion of the normal adult population does not evidence formal reasoning (Arlin, 1975; Kuhn, Langer, Kohlberg, & Haan, 1972; McKinnon & Renner, 1971; Neimark, 1975; Pitt, 1976; Tomlinson-Keasey, 1972, 1974; Schwebel, 1972). Might differences in reflective judgment level be attributable to level of formal operational reasoning?

Verbal fluency and socio-economic status are of interest because reflective judgment level is assessed by means of an interview and interview formats have been criticized for their susceptibility to the influence of verbal fluency and socio-economic status (Entwisle, 1972; Jensen, 1959; Kurtines & Greif, 1974). This study investigated whether higher reflective judgment levels were merely reflections of these two factors.

METHOD

Subjects

The sample for this study consisted of 60 subjects: 20 high school juniors, 20 college juniors majoring in liberal arts fields, and 20 doctoral level graduate students in liberal arts fields. Half of each group were males and half were females. The college juniors and graduate students were in attendance at a large
midwestern university at the time of the study, and the high school juniors attended schools in nearby communities. The high school and college groups were matched to the graduate group on sex, verbal scholastic aptitude when in high school, and size of home town when in high school. A graduate sample with the desired characteristics (doctoral level liberal arts majors with available high school scholastic aptitude scores) was identified by telephoning students whose names were randomly selected from the student-staff directory, placing an advertisement in the university student paper and taking recommendations from subjects who had already qualified for and had agreed to participate in the study. They ranged in age from 24 to 34, $M = 28.2$. The college sample was selected from a listing of all 19- to 20-year-old liberal arts juniors. The mean age was 19.6 years. The high school sample was selected from two high schools, one in an urban setting and one in a small town. They were selected from all 16- to 17-year-old juniors enrolled at the two schools who fulfilled the matching criteria. When more than one student had the desired matching characteristics, the subject was selected randomly. The mean age was 16.2 years.

The sampling procedures employed here have two major drawbacks: (1) age and educational level are confounded and (2) matching the level of verbal ability of the high school and college students with that of the graduate students is very likely to yield a sample of unusually bright high school and college students. The rationale for using these procedures was as follows. One of the most commonly stated goals of institutions of higher education is that students learn to reason complexly and with increasing adequacy about controversial issues. It seemed reasonable to assume that the more advanced levels of such reasoning would be found among graduate students, who, by merit of their training, had had greater exposure to educational experiences designed to fulfill this goal. In this way, the graduate students served as an "expert" sample. It also seemed reasonable to assume that high school students would have much less facility in reasoning about such issues, and that the college students would score between the other two groups. The age/educational level confounding was tolerated at this point in the research program in order to increase the probability of finding subjects in the higher positions.

Matching on verbal aptitude was employed as a means of controlling for the possibility that high reflective judgment scores merely reflect high verbal aptitude. In fact, matching the younger two groups with the graduate group on this variable yielded samples unrepresentative of high school or college students in general. For example, 75% of each student group scored above the 80th percentile on their high school scholastic aptitude test; the remainder scored between the 40th and 79th percentiles. Thus, the groups are well above average on this measure. However, a more generalizable procedure such as random selection within groups would be subject to the rather serious charge of having a much brighter graduate sample, thus potentially accounting for the hypothesized higher reflective judgment scores among the graduate students. In light of these con-
cerns, the strengths of this matching procedure across the three student groups were judged to outweigh the weaknesses.

Measures

Each subject was tested on five variables, described below.

**Reflective Judgment.** This is assessed by means of the Reflective Judgment Interview, which consists of a set of four dilemmas and a set of standardized probe questions presented by a trained interviewer in a semi-structured interview format. (It is a revised and shortened version of the interview used in the pilot study.) Each dilemma is defined by two conflicting and contradictory points of view on an issue, and subjects are asked to state and justify their points of view about the issue. The four dilemmas were chosen to represent different intellectual domains: science, current events, religion and history. (See King, 1977, and Kitchener, 1978, for a complete rationale for choice of dilemmas.) An example of one of the dilemmas follows:

There have been frequent reports about the relationship between chemicals that are added to food and the safety of those foods. Some studies indicate that such chemicals can cause cancer, making these foods unsafe to eat. Other studies, however, show that chemical additives are not harmful, and actually make the foods containing them more safe to eat.

After each dilemma is read aloud, a standardized series of probe questions is used to elicit subjects' rationale, for their points of view on the issue. These questions are designed to elicit statements regarding their view about the nature of knowledge, of evidence and of the role of authorities in resolving a controversy as well as their openness to alternative views or frames of reference. The order of presentation of the four dilemmas is randomized across subjects. The entire interview lasts 45 to 60 minutes.

In scoring the interviews, two trained raters independently score each subject's response to each dilemma. Since it is not assumed that subjects' responses are stage pure [see Rest's (1979) critique of the simple stage approach], scores from 1 through 7 were assigned to indicate the dominant reflective judgment stage evident in the protocol. A second score was assigned to indicate the sub-dominant stage. For example, a dominant stage 3/sub-dominant stage 2 protocol is scored as "332." An overall score represents the mean of all scores assigned by both raters. (See Kitchener, 1978, for a complete discussion of the issue of the assignment of scores in adjacent and nonadjacent stages.)

**Verbal Ability.** Although subjects were matched across groups on verbal scholastic aptitude, the Concept Mastery Test (CMT) was used as a measure of the current level of verbal ability, as a control for age/educational development on this variable. The Concept Mastery Test was designed by Terman (1973) to
measure abstract thinking at a high level. The test is comprised of two parts: the identification of synonyms and antonyms and the completion of analogies. The items were designed to draw on concepts which span a wide variety of fields. A high score usually indicates that a person has a large vocabulary as well as the ability to reason well using that vocabulary. The test is designed for group administration, takes about forty minutes to complete, and is objectively scored. It has been used with college and graduate students (Terman, 1973) as well as with bright high school students (Kennedy & Smith, 1963; Welsh, 1966, 1969).

**Formal Operations.** Subjects' level of formal operational thinking was measured by two of Inhelder and Piaget's (1958) tasks—the Combination of Colored and Colorless Chemical Bodies and the Oscillation of the Pendulum. These two tasks each reflect major components of formal operations, the combinatorial operations and the ability to separate variables. They have been found to load on a single factor (Bart, 1971; Lovell & Shields, 1967) and to be of equivalent difficulty (Lovell, 1961). These tasks take 15 to 30 minutes to complete.

Rating rules for the chemicals task were based on rules developed by Dale (1970) and elaborated by Pitt (1976) and King (1977). Mecke and Mecke's (1971) scoring system was used for the pendulum task. For each task, the possible scores ranged from 0-4.

**Socio-economic Status.** Demographic information, including a checklist of parents' occupation and educational level, was collected from each subject. Socio-economic status was computed using Hollingshead's (1957) two-factor index of social position.

**Verbal Fluency.** This variable was measured by counting the number of words spoken by the subject during the Reflective Judgment Interview. Word count was computed using the transcripts of each interview.

**Procedure**

Testing was conducted in two sessions. The Reflective Judgment Interview was administered in one session and the CMT and formal operations tasks were administered in balanced order in another session. Half the subjects in each student group received the CMT and formal operations tasks first; the remainder received the Reflective Judgment Interview first. Demographic information was collected at the beginning of each subject's first session.

**Reflective Judgment Interview.** Reflective Judgment Interviews were conducted by the first author of the study and a male colleague. Each interviewed one-half the subjects in each of the three student groups.
A student’s response to each dilemma was recorded on a separate audio tape. All tapes were coded to mask subjects’ identity and age; cues to age and educational level were deleted; and tapes were transcribed. The 240 separate dilemmas were rated independently by the authors according to the Reflective Judgment Scoring Manual. A mean score by dilemma was derived for each subject by averaging the scores assigned by both raters. An overall mean score for each subject was based in all ratings assigned for that subject across all dilemmas.

**Concept Mastery Test.** This test was administered individually and scored according to instructions in the test manual (Terman, 1973).

**Formal Operations.** Formal operations interviews were administered to the subjects individually in a balanced random order and followed the procedures described by Inhelder and Piaget (1958). These procedures are described fully by King (1977). The second author of the study and a trained, female colleague each conducted half the interviews for each student group. The experimental session for each task was tape recorded and the interviewer indicated on a tally sheet all trials conducted by the subject.

**RESULTS**

**Reflective Judgment**

**Interrater Reliability and Agreement.** Using a Pearson product-moment correlation, the reliability over all dilemmas for two judges rating independently was .96; for each dilemma, the reliability ranged from .86 to .90. Overall interrater agreement, the number of times two judges’ scores were discrepant by less than one stage, was 77%. Using Lawlis and Lu’s (1972) and Tinsley and Weiss’ (1975) adaptations of Cohen’s kappa (1960) to correct for chance agreement, the coefficient of agreement was 70%. Agreement levels within dilemmas ranged from 75% to 80%. No significant differences were found in reflective judgment scores when analyzed by raters or by interviewers over all dilemmas.

**Consistency across Dilemmas.** Consistency of scores across the four dilemmas within the Reflective Judgment Interview was evaluated by several means. Pearson product-moment correlations between all pairs of dilemmas ranged from .84 to .88, p < .001. Corrected item-total correlations (comparing subjects’ scores for one controversy with their scores on the other three) ranged from .89 to .92, p < .001. Internal consistency reliability (using Cronbach’s alpha) was .96. In addition to these standard indices, a frequency distribution of the scores assigned by both raters over all dilemmas was constructed for each
subject. This distribution indicated that the scores of 67% of the subjects spanned three stages or less, and only 8% of the subjects had scores that spanned more than four stages. Averaging over all subjects, 58% of a subject's ratings fell in his/her modal stage. (See Kitchener, 1978, Appendices O, P, and Q for complete data.) These results indicate that subjects evidence moderate to highly consistent scores within and across the four dilemmas.

**Scores by Student Group.** The major question addressed in the present study was whether students from three age/educational levels would reason at different stages on the reflective judgment model. A consistent upward progression of scores was found over the three student groups (see Table 1). Ninety percent of the high school students scored between 2.0 and 3.0, 85% of the college students scored between 3.0 and 4.5, and 70% of the graduate students scored above stage 5. The respective group mean scores were 2.77, 3.65, 5.67. The difference in mean scores across the three student groups was highly significant, \( F(2,54) = 76.44, p < .001 \). Differences among group means were tested using Tukey's-HSD Procedure. All pairs were found to be significantly different, \( p < .05 \). There was no effect due to sex or due to the interaction of group and sex. These results support the hypothesis that older subjects with more education reason at higher stages of reflective judgment than do their younger counterparts with less education. This finding leads to consideration of the second major question in this study: To what might these group differences be attributed?

**TABLE 1**

Distribution of Mean Reflective Judgment Scores by Student Group

<table>
<thead>
<tr>
<th>Mean Reflective Judgment Scores</th>
<th>High School</th>
<th>College</th>
<th>Graduate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4.5</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>6.0</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>6.5</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

| \( N = \) | 20 | 20 | 20 |
| \( M = \) | 2.77 | 3.65 | 5.67 |
| \( SD = \) | .49 | .81 | .92 |
| Range = | 2.17-4.04 | 2.46-5.58 | 3.83-6.92 |
Verbal Ability

Despite matching the three groups on high school scholastic aptitude test scores, there was a significant main effect for student group on the CMT, $F(2,54) = 35.02, p < .001$. (These scores appear in Table 2.) Tukey's-HSD procedure was used to test for all pair-wise differences between group means. All were significant in the expected direction, $p < .05$. Sex differences and the interaction between group and sex were not significant.

<table>
<thead>
<tr>
<th>Student Group</th>
<th>Reflective Judgment</th>
<th>Concept Mastery Test</th>
<th>Chemicals</th>
<th>Pendulum</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.77</td>
<td>40.75</td>
<td>3.55</td>
<td>3.00</td>
<td>740.43</td>
</tr>
<tr>
<td>SD</td>
<td>.49</td>
<td>17.72</td>
<td>.74</td>
<td>1.05</td>
<td>315.38</td>
</tr>
<tr>
<td>College</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.65</td>
<td>84.65</td>
<td>3.35</td>
<td>2.80</td>
<td>880.49</td>
</tr>
<tr>
<td>SD</td>
<td>.81</td>
<td>26.66</td>
<td>.85</td>
<td>1.23</td>
<td>322.95</td>
</tr>
<tr>
<td>Graduate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>5.67</td>
<td>109.80</td>
<td>3.30</td>
<td>3.00</td>
<td>1006.07</td>
</tr>
<tr>
<td>SD</td>
<td>.92</td>
<td>33.20</td>
<td>.84</td>
<td>1.10</td>
<td>258.37</td>
</tr>
</tbody>
</table>

The correlation between reflective judgment and verbal ability over all subjects was $.79, p < .001$. Within-group correlations between verbal ability and reflective judgment were smaller: $.63 (p < .001), .30 (p < .10) and .79 (p < .001) for the high school juniors, college juniors, and graduate students, respectively. Figure 1 shows within subject scores for both measures.

An analysis of covariance was run with group and sex as independent variables, reflective judgment mean score as the dependent variable, and verbal ability (CMT score) as the covariate. There was a strong significant main effect for student group on reflective judgment, $F(2,53) = 29.98, p < .001$, and no effect due to sex. Using Tukey's-HSD procedure, differences in adjusted reflective judgment mean scores between the graduate students and the other two groups were significant, $p < .01$ (see Table 3). Differences between adjusted mean scores of the high school juniors and college juniors were nonsignificant. Therefore, while reflective judgment and verbal ability are closely related, the differences between groups on reflective judgment cannot be solely attributed to level of verbal ability. It will be the task of future researchers to determine whether the strong relationship between verbal ability and reflective judgment
<table>
<thead>
<tr>
<th>CMT Score</th>
<th>0-20</th>
<th>21-40</th>
<th>41-60</th>
<th>61-80</th>
<th>81-100</th>
<th>101-120</th>
<th>121-140</th>
<th>141-160</th>
<th>161-180</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>△△△△</td>
<td>△</td>
<td>△△</td>
<td>△</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
<td>△△△△</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

△ H.S. Student
■ College Student
○ Graduate Student

Figure 1. CMT Score by Rounded Reflective Judgment mean score.
reflects the existence of a single underlying trait, or whether this reflects the existence of two distinguishable traits that covary for given populations. Given the nonsignificant correlation between these two measures obtained here for the college sample, this question seems especially important to pursue.

**Formal Operations**

Using Pearson's product-moment correlation, interrater reliability for two independent judges was .99 for the pendulum task and .98 for the chemicals task ($p < .001$). Interrater agreement, computed by counting the number of times both raters assigned the same score to a given protocol, was 97% for the chemicals task and 98% for the pendulum task. The mean scores of subjects tested by one interviewer were not statistically different from those of subjects tested by the other interviewer on either of the two tasks.

No differences between group means were significant for either the chemicals task or the pendulum task (see Table 2). Sex differences were not significant for either task, nor were the effects due to the interaction of group and sex. The most striking result of the formal operations data was the high proportion of subjects whose scores were either transitional or fully formal. Forty-three percent of the subjects received a transitional or a formal score on one of the two tasks and an additional 48% received a transitional or formal score on both tasks. Thus, fully 91% of the sample received a formal score on either one or both tasks, a finding that may be attributed to the overall high level of verbal ability of the subjects in this study. (Keating, 1975, for example, found a similar relationship between intelligence and formal operations.) This ceiling effect may account for the very low correlations found between reflective judgment level and formal operations scores, $r < .01$ for each task. In light of these results, no further analyses were run on this variable and little was revealed about the role that hypothetico-deductive reasoning plays in the ability to make reflective judgments. However, despite the attainment of formal reasoning abilities by almost all subjects in each group, highly significant differences in reflective judgment level were still found across groups. Therefore these differences cannot be attributed to differences in formal operational reasoning abilities.

**Verbal Fluency**

A count of the total number of words spoken in response to each reflective judgment dilemma was used as a measure of verbal fluency. A mean word count for each subject was derived by averaging the number of words spoken over the four dilemmas (see Table 2).

Using a two-way analysis of variance, a significant main effect by group was found, for word count, $F(2,54) = 3.58, p = .03$, but not for sex or for the interaction between group and sex. Using Tukey's-HSD procedure, the graduate
students' mean was significantly higher than the means of the other two groups, \( p < .05 \). The difference between the mean of the high school group and college group was not significant. The overall correlation between reflective judgment and word count was \( .40, p < .001 \). An analysis of covariance was used to test for differences between groups on reflective judgment while controlling for word count. Overall differences between groups on reflective judgment remained significant, \( F(2,53) = 62.34, p < .001 \), as were all the pair-wise comparisons of adjusted mean scores using Tukey's-HSD procedure, \( p < .05 \) (see Table 3). Sex differences and the interaction between group and sex remained nonsignificant. Therefore, while verbal fluency correlates moderately with reflective judgment, the differences between groups on reflective judgment cannot be solely attributed to verbal fluency.

**Socio-economic Status**

Socio-economic status correlated weakly but significantly, \( r = .22, p < .05 \), with overall reflective judgment score. An analysis of covariance was used to test for group by sex differences in reflective judgment while holding socio-economic status constant. Differences between groups remained significant, \( F(2,53) = 70.12, p < .001 \). Sex differences and the interaction between group and sex were not significant. Differences between adjusted group means were tested using Tukey's-HSD procedure. All pairs were significantly different, \( p < .01 \). In other words, social status did not statistically account for the differences between groups found in reflective judgment scores.

**Combining Covariates**

Verbal Ability and Verbal Fluency. An analysis of covariance was used to test for group differences in reflective judgment while holding both verbal ability (CMT score) and verbal fluency (word count) constant. A significant main effect was found for group, \( F(2,52) = 27.13, p < .001 \), but not for sex or for the interaction between group and sex. Adjusted mean scores were almost identical to those adjusted for verbal reasoning alone (Table 3).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>CMT, Word Count</th>
<th>CMT and SES</th>
<th>CMT, Word Count, and SES</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School</td>
<td>3.42 2.78 2.80</td>
<td>3.45 3.48</td>
<td>3.45</td>
</tr>
<tr>
<td>College</td>
<td>3.52 3.65 3.63</td>
<td>3.53 3.52</td>
<td>3.52</td>
</tr>
<tr>
<td>Graduate</td>
<td>5.13 5.66 5.66</td>
<td>5.11 5.06</td>
<td>5.12</td>
</tr>
</tbody>
</table>
**Verbal Ability and Socio-economic Status.** An analysis of covariance was used to test for overall differences in reflective judgment while controlling for socio-economic status and verbal ability. Differences between groups were significant, $F(2,52) = 29.36, p < .001$, whereas no effect was found for sex or for the interaction between group and sex. Adjusted mean scores fell very close to those adjusted for verbal ability alone (see Table 3). Differences between group means were tested using Tukey’s-HSD procedure. Differences in the means of the graduate students and the means of the other groups were significant, $p < .01$, whereas the difference between the means of the high school group and the college group was nonsignificant.

**Verbal Ability, Verbal Fluency and Socio-economic Status.** These analyses indicate that controlling for verbal fluency and socio-economic status added little to the effects of controlling for verbal ability alone in this sample. As when controlling just for verbal ability, the differences in adjusted reflective judgment scores are significant between the graduate students and each of the other two groups, $p < .01$, but are not significant between the high school and college groups. This difference remained even after controlling for the effects of verbal aptitude, verbal fluency, and socio-economic status.

**DISCUSSION**

In this study, reflective judgment scores were found to differ across three age/educational levels. Older, better-educated subjects held more complex and sophisticated assumptions about the justification of beliefs than did younger subjects with less education. The 16- to 17-year-old high school students tended to justify their beliefs with absolute assumptions about knowledge or assumed there was no valid way to justify beliefs unless knowledge was manifest in data. By contrast, the 19- to 20-year-old college students tended to assume that absolute knowledge was for practical reasons impossible to attain, and they tended to view the justification process as idiosyncratic. Graduate students (aged 24 to 34) tended to assume that knowledge claims needed to be and could be rationally justified as reasonable conjectures about reality. These differences in reflective judgment level across groups could not be statistically accounted for by other theoretical or potentially confounding factors related to intellectual development: verbal ability, formal operations, socio-economic status, and verbal fluency. Whereas verbal ability was found to be closely related to reflective judgment, the differences between groups on reflective judgment could not be solely attributed to this factor; for example, the low correlation between reflective judgment and verbal aptitude in the college group suggests that the development of reflective judgment and verbal aptitude are not parallel processes. It may be that certain levels of verbal aptitude are necessary but not sufficient for higher levels of reflective judgment, but that different experiences stimulate the development of
each. It may also be that reflective judgment and verbal ability are two distinct traits that happen to covary among the population tested.

To what, then, might the differences in reflective judgment level be attributed? Given the sampling limitations noted above, differences across these groups may reflect differences in maturation, education, selection into higher educational programs, or these factors in combination (e.g., an interaction between maturation and education). The selection factor is particularly relevant for the graduate sample in that being admitted to and successfully completing a graduate program may partially be a function of valuing intellectual dilemmas and of being able to make convincing arguments about such dilemmas.

The implications of these findings are important for educational institutions which are increasingly being challenged to demonstrate the effectiveness of their programs for students, not just in quantitative terms such as gaining more knowledge, but also in qualitative terms such as reasoning more coherently and persuasively. This work in the development of reflective judgment offers instructors information about how students at different age/educational levels differ in the ways in which they justify their beliefs and suggests that these are related to their basic assumptions about knowledge and reality. In addition, the model suggests that these beliefs are not just acquired through learning, but that higher forms of justification develop out of lower forms in a sequential manner. This research suggests that development of reflective judgment probably occurs through an interaction of environmental and maturational factors.

To the extent that a major goal of higher education (and especially of graduate training) is to teach students how to be mature inquirers who participate in the construction and critique of theory and knowledge and to the extent that this goal may be understood as an outcome of a developmental sequence, then a description of the sequence is an essential prerequisite to understanding how this goal can be achieved. For example, the degree to which students' assumptions about reality and knowledge color how they understand the learning process and consequently education itself may account for the varied interpretations of the educational process which students hold and which are often in conflict with the viewpoints of their instructors.

These findings also have important implications for understanding how mature adults make judgments about complex problems. In making decisions about current issues, whether in a scientific, political, or current events area, the individual is often confronted with conflicting data, opinions, and theories. Evaluating and making judgments about complex issues is partially dependent on the individual's initial assumptions. To reason critically means to examine different points of view, reflect on their logic, evaluate the evidence, and come to a conclusion about what seems most reasonable or likely. High levels of scientific and philosophical creativity are dependent on such critical processes and preclude the assumption of absolute authority-based knowledge. If this model offers an accurate description of concepts of justification and their development, then this line of research clearly warrants further study and elaboration.
REFERENCES


Gilligan, C., & Murphy, J. M. Development from adolescence to adulthood. The philosopher and the dilemma of the fact; New Directions for Child Development: Intellectual Development Beyond Childhood, 1979, 5, 85–99.


Kitchener, K. S., & Kitchener, R. The development of natural rationality: Can formal operations


Meeke, G., & Meeke, V. The development of formal thought as shown by explanations of the oscillation of pendulum: A replication study. Adolescence, 1971, 6, 219-228.


Schwebel, M. Logical thinking in college freshmen. Final report, Project No. 0-B-0105, Grant No. OEG-2-7-0039 (509), April, 1972.


Tomlinson-Keasey, C. Formal operations in females from eleven to fifty-four years of age. Developmental Psychology, 1972, 6, 364.


