# What Constitutes Skilled Argumentation and How Does it Develop?

MARION GOLDSTEIN
AMANDA CROWELL
DEANNA KUHN

*Teachers College*
*Columbia University*
*525 West 120th St.*
*New York, NY 10027*
*U.S.A.*

*mariongoldstein@gmail.com*
*amanda.jane.crowell@gmail.com*
*dk100@columbia.edu*

**Abstract:** We report our efforts to assess the skill of contemplating and evaluating argumentation. An adaptive forced-choice instrument was developed and administered to 6th grade students, 7th grade students who had participated in a year-long intervention that successfully strengthened their argumentation production skills, and expert arguers. The instrument was sensitive enough to detect differences in skill level across these groups. Despite their gains in production skill, however, 7th graders showed only modest superiority over the untrained 6th graders and performance well below that of experts. Meta-level demands involved in evaluation but not present in production, we propose, may make evaluation more difficult.

**Résumé:** Nous faisons un rapport sur nos efforts d'évaluer les habiletés d'interpréter et d'évaluer l'argument-ation. Nous avons créé une épreuve qui mesure ces habiletés et l'avons administrée à trois groupes différents: des étudiants de 6ième année, des étudiants de 7ième année qui ont réus-sis à améliorer leurs habiletés argu-mentatives durant une année d'inter-ventions, et des experts. L'épreuve était assez judicieuse pour détecter les différents niveaux de performances dans ces trois groupes. Malgré l'amé-lioration des étudiants de 7ième année, leurs résultats de l'épreuve étaient seulement modestement supérieurs à ceux des étudiants non entraînés de 6ième année, et très inférieurs à ceux des experts. Nous proposons que les exigences à méta-niveaux impliquées dans l'évaluation peuvent rendre l'évaluation plus difficile.

**Keywords:** argument; argumentation; development; metacognition; causal reasoning

## 1. Reasoning as argumentation

Cognitive psychologists have a long-standing interest in reasoning, as well as more elemental cognitive processes, and believe the reasoning of ordinary individuals to be amenable to empirical investigation. Yet, until recently, virtually all of their research attention in the realm of reasoning has been devoted to solitary problem solving by individuals, most often of well-structured problems of an artificial nature, such as the extensively studied Tower of Hanoi problem. Argumentation, in contrast, defined as asserting and defending claims consistent with one's purposes, has a social dimension and is ubiquitous in everyday activity. Indeed, the case has been made by Oaksford, Chater, and Hahn (2008) that argumentation is the umbrella under which all reasoning lies; in their words, it is "the more general human process of which more specific forms of reasoning are a part" (p. 383). Asserting, supporting, and refuting claims is the purpose to which we apply our reasoning skills. If so, the repeatedly noted poor performance of students in assessments of argumentation skill (see Kuhn & Franklin, 2006; Kuhn, in press, for review) becomes a cause for serious concern.

## 2. Assessing skills of argument

We describe here our efforts to identify and measure argumentation skills in adolescents and adults. In particular, we focus in this article on evaluative skills. There have been some studies of people's generally weak skills in the evaluation of individual arguments, but little research that we know of on meta-level, or evaluative, skills with respect to dialogic argumentation. In earlier work we describe here, we have seen students improve over time in the practice of dialogic argumentation. Do they, we wondered, show corresponding developmental change in the appreciation of stronger vs. weaker argumentive moves during discourse? Increasing use of stronger moves implies such awareness, but we thought the question was worthy of empirical test.

We begin by noting similarities and differences between our general approach and those of several others engaged in research on argument. The first distinction to be made is that between argument as a product and argumentation as a process. Studies of the former focus on rhetorical arguments in support of a claim advanced by a single individual in the absence of an interpersonal or dialogic context. Analyses of such arguments focus on formal structure and the presence or absence of individual components regarded as necessary to sound arguments. As noted by Clark, Sampson, Weinberger, and Erkens (2007) in their review of approaches, most researchers adopt a framework that draws on

Toulmin's (1958) analysis of the components of argument. Toulmin proposed that an argument consists of six possible components. These include claims (the conclusions drawn by the arguer), data (statements used as evidence to support a claim), warrants (statements that relate the claim to data), qualifiers (special conditions that present the arguer's degree of certainty about the claim), backings (underlying assumptions that provide justifications for the warrant), and rebuttals (statements that acknowledge the limits of a claim). Subsequent researchers have simplified Toulmin's framework (e.g., Erduran, Simon, & Osborne, 2004), collapsing categories in order to improve clarity and reliability of analysis.

Regardless of the adaptations, stronger arguments are thought to contain more of the different components than weaker arguments. The accuracy or relevance of the statements within an argument is accorded lesser importance in determining an argument's strength. Researchers who use a framework focusing on formal argumentation structure are able to identify the argument components that students tend to omit during argument construction. Pedagogical supports may then aim to address these skill deficits (Larson, Britt, & Kurby, in press).

Other researchers have focused on types of reasoning individuals use in an argument, more than the formal structure of the argument. These approaches are concerned with the quality of the grounds and the accuracy of the content within an argument. Arguments that distort evidence or address irrelevant points, for example, are considered weaker than arguments that are accurate and logically coherent. Jimenez-Aleixandre, for example, examines students' treatment of claims by classifying their reasons into epistemic categories including inductional, deductional, appeals to analogy, and others (Jimenez-Aleixandre & Erduran, 2008). Similarly, Duschl (2008) has suggested a system originating with Walton (1996), in which students' arguments are classified in terms of requests for information, expert opinion, inference, and analogy. In each of these cases, the proportions of arguments in each category typically are tracked to identify changes in the nature of students' reasoning over time.

We turn now to our focus here, analyses of argumentation as a dialogic process in which two or more individuals engage. Several researchers have studied this dialogic process and undertaken to identify its characteristics, some at the more macro level of an entire dialog or dialogic sequence and others, like our own, at the micro level of individual utterances. Leitäo (2000, 2003), for example, analyzes dialogic sequences and posits that the most successful argumentive interactions adhere to a specific pattern involving a claim, a responsive counter-claim, and an integrative reply that incorporates the previous ideas. Strong arguments,

according to this framework, build on the ideas of participants and, over time, differences in perspectives are negotiated and resolved. Erduran et al. (2004) classify conversational turns into a hierarchical category system ranging from simple exchange of claims to single and multiple rebuttals. Clark and Sampson (2008) similarly propose a category system that includes analysis of discourse moves as well as conceptual quality of arguments.


## 3. A functional coding scheme for dyadic argumentation

Our own efforts to develop a system of analysis applicable to dialogic argumentation (Felton & Kuhn, 2001; Kuhn & Udell, 2003, 2007) were based on consideration of the goals of argumentive discourse. Specifically, we drew on the framework originating with Walton (1989), who posits two goals of argumentation: a) to obtain commitments from the opponent to support one's own position, and b) to challenge and weaken the opponent's position by critiquing their premises. Both of these goals mandate attention to the ideas of the opposing side. In this way, the framework is transactional in nature because the strength of the argument, or whether an argument is deemed productive, is determined by whether and how the arguer addresses the ideas put forth by the interlocutor.

In our functional coding scheme (Felton & Kuhn, 2001; Kuhn, Goh, Iordanou, & Shaenfield, 2008; Kuhn & Crowell, in press), a code is assigned to each utterance in a dialog. These codes are based on an utterance's functional relation to the opponent's preceding utterance. Specifically, what *function* does the utterance serve in relation to the immediately preceding utterance and with respect to the mutual goals that define the objective of the interchange? The weakest arguments simply express agreement or disagreement without mentioning a rationale, or express one's own ideas in the absence of any attention to the other side. At the next level, the arguer goes beyond simple disagreement to advance an argument to justify this disagreement. Here, distinction between two different forms of counterargument becomes critical. The weaker form of counterargument, a Counter-Alternative, expresses disagreement by advancing a different argument in support of one's own position but does not directly address the argument put forth by the opponent, thereby leaving its force intact. For example, in a dialog in which 6th grade students were asked to argue about the topic "Should a misbehaving student be expelled from school?," one boy was confronted with his opponent's statement, "They shouldn't be expelled because they deserve another chance." His reply ignored his peer's "another chance" concept and instead introduced a new, perfectly sound argument against the opponent's

*position*, but one that leaves the opponent's *argument* intact: "Yes but they have been acting up for a while and their behavior has not gotten better and it's not fair to the other kids who are trying to learn."

A Counter-Critique, the stronger of the two counterargument types, disagrees by responding directly to the opponent's argument with an argument designed to weaken its force. For example, in a dialog later in the year, the 6th graders were arguing about another topic (Should the sale of human organs be allowed?), and the same boy exhibited greater skill in genuine counterargument. His opponent argued, "[They] shouldn't be allowed to [sell their kidney] because it is part of their body," and this time he directly addressed and undertook to weaken his opponent's argument: "But if people are willing to give up their own body parts and be so generous to the people who need kidneys why should we stop them…?" Over time and with extended opportunity to practice, students' transition from the frequent use of weaker argumentive strategies to more frequent use of powerful strategies, notably Counter-Critique, reflects development in the production skills central to argumentive discourse.

Skill in dialogic argumentation is clearly significant in its own right—a skill that pays dividends not only in academic contexts but in the world of work and in everyday life. In our work on developing argument skills in adolescents (see Kuhn & Crowell, in press, for a current review), we have focused on dialogic argumentation not only for this reason but also because we, along with a number of others (Billig, 1987; Graff, 2003) see it as a promising pedagogical path to the development of individual expository argument skills in both verbal and written forms. Our recent work (Kuhn et al., 2008) provides some initial support for this view. In an extended, year-long intervention, middle-school students engaged in electronic dialogs about a series of social issues such as whether a misbehaving student should be expelled from school, whether parents have the right to homeschool their child, and whether the USA should intervene to help a small third-world country under attack. Students were assigned to positions based on their responses to an opinion poll and, working with a same-side partner, were asked to convince a succession of peers holding an opposing view that their position was the better one. Electronic dialogs lasted about 30 minutes, and each topic was debated 5-6 times over the course of several months, each time with a new pair from the opposing side. This activity culminated in a final, full-class debate. Over the year, students demonstrated gains in what we have referred to as argument production—showing more frequent use of the more skilled strategy of counterargument (the Counter-Critique) in their dialogs. They also showed gains in individual (non-dialogic) argumentive essays.

A third aspect of skilled argument, in addition to skill in argumentive discourse and in production of individual expository arguments, is skill in argument evaluation. We turn now to describe our recent efforts to investigate this third component. Both academic and research assessments have shown students to be weak in evaluation, as well as production, of individual expository arguments, but little is known about the skills involved in the evaluation of dialogic argument. In our dialogic-based interventions to develop argumentation production skills, such as the one described above, we have had success in enhancing performance of both dialogic and individual production of sound arguments. How might one assess skill in evaluation of dialogic argument, we wondered, and would we see corresponding advance in evaluation skill following interventions that have been successful in enhancing the two performance components?

## 4. Assessing skill in evaluation of dialogic argument

In the design of such an instrument, we focused our attention on what our work on production of dialogic argumentation has suggested is a key evolution—the transition illustrated above from less advanced levels to responding to an opponent's argument by identifying its weaknesses and thereby reducing its force (the Counter-Critique). Prior to this transition, our study of dialogic production has shown (see Kuhn & Crowell for review), arguers initially ignore the opponent's argument entirely and focus exclusively on their own arguments supporting their own opposing claim; or, if they have begun to pay some attention to the opponent's argument, they express their disagreement with it, supported by a new argument against the opponent's position that leaves the opponent's immediately preceding argument unaddressed (the Counter-Alternative).

We developed an instrument to measure students' ability to distinguish between these two argumentive strategies—the Counter-Critique and the Counter-Alternative. It situates students in a hypothetical argument with a friend. This friend, who we call Lee, begins with a claim about why some students fail in school. Upon reading Lee's argument, the student is presented with two options for how to respond to Lee. One choice is a strong counterargument—a Counter-Critique—and one is a weak counterargument—a Counter-Alternative. Students are asked to imagine that they disagree with Lee's position and choose the response to Lee that is in their opinion the stronger of the two. The response the student chooses dictates Lee's rebuttal, which is a Counter-Critique response to the reply selected by the student. Again, the student has two options for how to respond to Lee's

rebuttal. In this way, the instrument is designed to mimic genuine argumentation and permits an assessment of the extent to which respondents are able to recognize and select stronger responses to their hypothetical opponent. When the respondent completes three counterargument selections relating to why students fail in school, the topic changes to why many criminals return to a life of crime after being released from jail. Again, for each of three items, Lee presents claims and the student must choose the better of two counterarguments. The third topic is whether movie stars should make as much money as they do. Pilot testing showed the three to be of comparable complexity and difficulty level. Response choices that respondents in the pilot sample found confusing or of low plausibility were revised or eliminated and replaced. All alternatives thus had at least surface plausibility as accurate statements with respect to the topic. To minimize the likelihood of respondents' choosing a response option based on factors other than its relevance to the preceding claim, we ensured that both options at each decision point were of comparable length and comprehensibility, as well as accuracy as true statements.

Although respondents dictate their own path through the assessment based on the response options they choose, each respondent follows a path in which he/she chooses three counterarguments for each topic. Thus, respondents are confronted with nine decisions over the course of the assessment. A respondent's score is the number of Counter-Critique replies he/she selected. Nine is the highest possible score.

The instrument is housed in SurveyMonkey, a Web-based tool used to create surveys and other self-report instruments. The tool's skip logic enables the designer to dictate which test item follows a respondent's selection. This allows the instrument to absorb respondents in an actual sustained dialogic argument by ensuring that they are presented relevant, factually plausible rebuttals to their choices of counterarguments. This format, we believe, is superior to independent, isolated test items in that it is more reflective of authentic discourse. An example of the problem structure for one of the three topics, why some students fail in school, is presented in Figure *1*. Respondents, however, never see the entire structure; instead, they see only the two response options at each point, determined by their preceding choice.

*Figure 1. Problem structure for one topic; left branches are the stronger response option (the Counter-Critique).*

## 5. Initial findings

Initial testing of this instrument was undertaken with 6th and 7th grade students (ages 11-13) at an urban public middle-school. The student body is 75% minority (primarily African American and Hispanic, with a small number of Asian students). Students are residents of the surrounding low- to middle-income neighborhood.

Beginning in 6th grade and continuing through 8th grade, students at this school participate in a twice-weekly philosophy class designed to foster the development of argumentation skills. Work with this curriculum (Kuhn & Udell, 2003; Kuhn et al., 2008) supports the view that argumentive discourse skills develop gradually through authentic practice in argumentation. As described earlier, the curriculum provides dense experience in argumentive discourse as students debate real-world social issues, first in interchanges with a succession of peers holding an opposing view and finally in a whole-class debate.

We administered the argumentation evaluation instrument described above to 92 of these 6th grade students at the beginning of the school year, prior to their exposure to the argument curriculum, in order to assess the evaluation skills of novice arguers of this age group. As explained earlier, a student's score is the number of Counter-Critique responses he/she selects. Students receive a 9 if they choose the stronger counterargument—representing a Counter-Critique—at each decision point. In this assessment, 6th grade students achieved a mean score of 5.66.

Because each item has only two response options (and therefore students have a 50% chance of choosing the Counter-Critique response), chance responding on all 9 items can be expected to lead to a score of 4 or 5. For this reason, we classified scores of 0-5 in a low performance category. Scores of 6 and 7 were assigned to a middle category, as possibly showing some competence in choice of the Counter-Critique option, and scores of 8 or 9 were classified as reflecting high performance and clear recognition of Counter-Critique as the superior alternative. Of 92 6th grade students who completed the assessment, 41 (44.57%) received low scores, 32 (34.78%) received middle scores, and 19 (20.65%) received high scores.

We also administered the instrument to 63 students at the beginning of their 7th grade year. These students had completed the first year of the three-year argument curriculum. The mean score for this group was 6.17. The direction of the difference between the 6th and 7th grade scores suggests some increasing evaluation skill with increasing age and experience, but this mean score is not significantly higher than that of the 6th grade students, $t(150.7)=-1.731$, $p=.10$. Hence, the increase is slight at best. Of the 63 7th grade students who completed the instrument, 20 (37.15%) received a low score, 28 (44.44%) received a middle score, and 15 (23.81%) received a high score. Future research will track students' skill levels throughout the course of the three-year curriculum. In this ongoing work, we have expanded the instrument to a five-topic, 15-item one to avoid ceiling effects.

In addition, as an expert comparison group, we administered the assessment to a group of 37 doctoral students in developmental or educational psychology. The mean score for experts was 7.46. Four experts (10.81%) received a low score, 13 (35.14%) received a middle score, and 20 (54.05%) received a high score.

Mean scores across the three groups differed significantly, $F(2,189)=12.56$, $p<.001$. This effect establishes that the instrument is sensitive enough to detect differences in argumentation evaluation skills. It also establishes that young adolescents lack proficiency in this area, with only at best modest gains and much room for improvement after a full year of dense practice in argumentation.

Results with respect to production tell a different story. Here practice leads to significant gains (Kuhn & Crowell, in press). In September and again in May, the students reported on above engaged in one-on-one dialogic argumentation on the topic of capital punishment, a topic the students had not discussed during the intervention. Students were assigned to Pro/Con sides based on their responses to an opinion poll, and each student argued against another student who held an opposing view.

Figure 2 shows students' performance across time on the capital punishment dialogic assessment, compared to a control group of the same grade and from the same school who participated in a parallel whole-class discussion class revolving around similar topics. Figure 2 shows mean percentage of utterances that were successful Counter-Critiques of the opponent's preceding utterance, in September and again in May. At the May posttest, 35% of the intervention group students responded with Counter-Critiques more than one third of the time, compared to 13% of control group students, a significant difference (Kuhn & Crowell, in press).
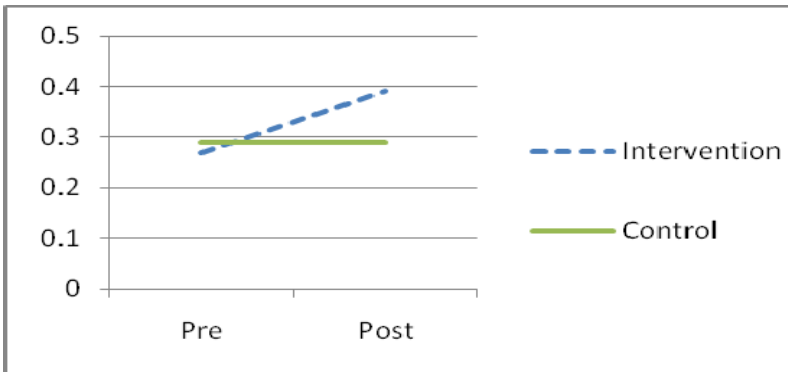


Figure 2. Production skill in dialogic argumentation at pre- and post-tests, indexed by percentage Counter-Critiques.

## 6. Interpreting the relation between evaluation and production of skilled counterarguments

The production results reported above, as well as our prior intervention studies (Kuhn & Udell, 2003; Kuhn et al., 2008), show that performance assessments of middle school students' dialogic argumentation improve following dense engagement in dyadic argumentation with peers. Our purpose in the work described here has been to develop a means to assess what we might expect to be a parallel skill—the *recognition* or appreciation of superior moves in dyadic argumentation. If so, we might then assess whether

corresponding improvement in recognition occurs with time and engagement.

If anything, we might predict that recognition skills would not simply parallel production skills but would in fact precede them, a pattern that has been noted in a broad range of domains from early language development to moral reasoning development: An individual first recognizes and can appreciate (and hence at least implicitly evaluate) a higher form prior to being able to independently produce it.

To the contrary, however, while the middle school students reported here on average show themselves capable of learning to produce the higher-level, Counter-Critique counterarguments, they do not show such improvement in learning to recognize the superiority of direct Counter-Critique when given a choice between it and a less advanced argument strategy. How should this pattern of performance be interpreted? This is the question we turn our attention to in the remainder of this article.

## 7. Why is argument evaluation challenging?

As a first step toward addressing why evaluation skills appear to be difficult to develop, we return to the simpler case in which individuals are asked to evaluate the strength of a simple (non-dialogic) argument in support of a claim. What challenges does this task present? In our own earlier work addressed to this question (Kuhn & Felton, 2000; Kuhn, 2001), we asked 8th graders and adults to choose the stronger of two arguments in support of a claim. One provided a theoretical explanation that made the claim plausible, whereas the other provided empirical evidence that the claim was true. Earlier work of our own and by others has indicated a preference for explanation over evidence as justification for a claim (Brem & Rips, 2000; Kuhn, 2001). The following is an example of the choice that our items presented:

Why do teenagers start smoking? Which is the stronger argument?

A.  Smith says it's because they see ads that make smoking look attractive. A good-looking guy in neat clothes with a cigarette in his mouth is someone you would like to be like.

B.  Jones says it's because they see ads that make smoking look attractive. When cigarette ads were banned from TV, smoking went down.

After choosing A or B as the stronger argument and providing a justification, respondents were also asked explicitly to indicate not

only the strengths of the argument they chose and the weaknesses of the other argument, but also whether the chosen argument had any weaknesses and the non-chosen argument any strengths.

A very common response, especially among the 8th grade respondents, was to cite nonepistemic, rather than epistemic, criteria as justifications for their choices and evaluations. Epistemic criteria apply to any argument of a given form. Nonepistemic criteria apply only to an argument of specific content. Nonepistemic justifications thus most often address the correctness of the claim ("This is a good argument because what it's saying is true"), rather than the quality of the argument supporting the claim. Less often did 8th graders (as well as a number of adults) invoke epistemic criteria, citing the epistemic strength of explanation (e.g., "It gives a reason"), or the epistemic strength of evidence (e.g., "It's something that really happened"). Even less often did teens or adults mention the epistemic weakness of explanation (e.g., "It's only a theory") or the epistemic weakness of evidence (e.g., "It doesn't say why").

These findings indicate that adolescents and even some adults find it difficult to think about the formal characteristics of an argument. In contrast, if the content and meaning of the claims involved are familiar, they find it easy to evaluate this content, at least with respect to their agreement or disagreement with it.

In this case, note, the respondent's own perspective becomes one with one or the other of the two argument proponents (Smith and Jones) whose arguments are being compared. The respondent "agrees" with what Smith (or Jones) is saying and adopts the argument as his or her own. To invoke an epistemic criterion, in contrast, the respondents must adopt a distanced, "third-party" stance in order to evaluate Smith's (or Jones') argument and its standing in relation to the claim at hand. Hence a meta-level (metacognitive) stance is required, to enable reasoning about reasoning.


## 8. Non-dialogic vs. dialogic argument evaluation

If we extend this analysis to the case of dialogic argumentation, the context of primary interest to us here, we can see that similar considerations apply. In the argumentation evaluation instrument described earlier, the respondent is asked not simply to evaluate the two choices on their own merit as arguments, but rather to evaluate them in relation to (and specifically as counterarguments to) the preceding argument. Although a seemingly straightforward task on the surface, the cognitive demands it poses are in fact considerable. The respondent likely holds views of agreement or disagreement with the content of the initial argument as well as each of the

proposed counterarguments. Each of these stances must be in effect bracketed—held in abeyance—to allow execution of the task: the evaluation of each of the proposed counterarguments with regard to its *relation* to the initial argument. As in the simpler non-dialogic case, a meta-level stance is required, but now in an even more complex form given the number of propositions involved.

Following administration of the dialogic argument evaluation instrument to the middle school students described earlier, we interviewed a number of them to gain insight into the thinking underlying their choices. The following response from a 7th grader is typical of what we heard. At the choice point being discussed, her task is to choose between A and B as the stronger counterargument to the argument preceding it.

Prisoners who get released from jail return to a life of crime because they can't find a good job.

A. Jails offer training to prisoners so that they can find work when they get out.

B. When prisoners get out of jail, their old friends often pressure them to return to crime.

This student chose B as the stronger counterargument and when asked to explain why, she said the following:

> Because I have family members; my dad's best friend has been in jail twice, so I know from personal experience. They're accustomed to crime so there's pressure and then they're used to it so they continue to commit crimes.

Clearly, she has ignored the initial argument and responded based only on her agreement with the content of the second counterargument. She is endorsing a particular causal explanation of a particular phenomenon, but she is not engaged in the evaluation of arguments or argumentation.

In this example, we in fact have no evidence of the student's processing of the first alternative. We don't know if she agrees or disagrees with it or if she has compared it to the second alternative.

In another example, a student does explicitly address both alternatives:

Students fail in school because they don't try hard enough to do well on tests.

A. No matter how hard students work, some just aren't good test-takers.

B. Some students act out in class instead of paying attention to the teacher.

This student chose B as the stronger counterargument and gave the following explanation:

> Because I don't think it's true. There is no such thing as good and bad test takers. But some students do play around and don't listen and just give up.

As in the first example, however, this student gives no indication of considering either alternative in its relation to the initial argument.

There is much evidence in the cognitive development literature of children and adolescents having difficulty reasoning about the formal relations among propositions independent of their truth value, in the case of both deductive and inductive reasoning (Kuhn & Franklin, 2006; Klaczynski, 2004; Moshman, 2005). In this light, the present indications of difficulty in argumentation evaluation should not be a great surprise.

A further contributor to this challenge may possibly lie in children's and adolescents' conceptions of causality, in particular the concept of multiple causes contributing to a common outcome. Students of middle-school age have considerable difficulty in coordinating the effects of multiple variables on an outcome (Kuhn, 2007; Kuhn, Pease, & Wirkala, 2009). They are susceptible to what has come to be known as discounting (Sedlak & Kurtz, 1981): identifying one cause of an outcome makes other causes less likely. To see how this tendency may figure in argumentation evaluation, consider the above example. Clearly, there exist multiple reasons that students fail in school. The "act out in class" cause should not foreclose the possibility, or even reduce the likelihood, of other causes. To the extent one does not think in terms of multiple causes, however, endorsement of one cause (as the student in the example does by choosing counterargument B) may do exactly this—reduce the perceived likelihood of other causes being relevant. Accordingly, they become less worthy of contemplation and of argument.

## 9. Conclusion: Reconciling production and evaluation

In conclusion, we return to our earlier question. After an extended and dense intervention affording engagement and practice in dialogic argumentation, young adolescents show significant improvement in their ability to address an opposing peer's arguments with cogent counterarguments that serve to weaken their

force. They do not, however, show much improvement in recognizing the stronger of two counterarguments with respect to their power to weaken a claim. We have suggested some of the cognitive challenges the latter poses. Confirming their role and more precise nature clearly requires further exploration. We believe such investigation is warranted, given the fundamental significance of argument and argumentation in everyday thinking, as well as its importance to the intellectual development of students in academic contexts, especially those who go on to advanced levels of education.

The qualification is in order that in the work described here we did not undertake to measure and compare production and comprehension skills on comparable instruments, equated for content and difficulty in all other respects. The conclusion we wish to draw is simply that an intervention focused on argumentation that produced significant advances in production skill did not yield corresponding advances in evaluation skill. This finding suggests a possibility that initially may seem counterintuitive; recognizing the relative strength of counterarguments may not precede the ability to produce them when engaged in one's own authentic argumentation. In the latter case, the meta-level demands are lesser. The other's argument opposes one's own, and the counterargument to be produced is one that coincides with one's own views. Furthermore, the motive to weaken the force of the other's argument to the greatest extent possible is clear, consistent, and strong. No "third-party" stance-taking is required. All of these factors help to make the task more tractable.

Moreover, and finally, the dialogic production task may serve as an effective bridge to the development of the recognition and evaluation skill that the students we have described seem to have found the most challenging. With continued engagement and practice in dialogic argument across multiple topics and opponents, the meta-level demands of evaluation may become less challenging. Alternatively, they may need to be met by activities that engage them directly. These are the alternatives we hope to assess in continuing work.

## References:

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to social psychology.* Cambridge: Cambridge University Press.

Brem, S., & Rips, L. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573-604.

Clark, D., & Sampson, V. (2008). Assessing dialogic argumenta- tion in online environments to relate structure, grounds, and

conceptual quality. *Journal of Research in Science Teaching, 45(3)*, 293-321.

Clark, D., Sampson, V., Weinberger, A, & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19(3),* 343-374.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32,* 268–291.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915-933.

Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind.* New Haven: Yale University Press.

Jiménez-Aleixandre, M.P., & Erduran, S. (2008). Argumentation in science education: An overview. In Erduran, S., & Jiménez-Aleixandre (Eds.). *Argumentation in science education: Perspectives from classroom-based research*, (pp. 137-157). Dordrecht: Springer.

Klaczynski, P. (2004). A dual-process model of adolescent development: Implications for decision making, reasoning, and identity. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 31). San Diego: Academic Press. (pp. 73-123).

Kuhn, D. (2001). How do people know? *Psychological Science, 12*, 1-8.

Kuhn, D. (2007). Reasoning about multiple variables: Control of variables is not the only challenge. *Science Education, 91,* 710-726.

Kuhn, D. (in press). Reasoning. In P. Zelazo (Ed.) *Oxford Handbook of Developmental Psychology.* New York: Oxford University Press.

Kuhn, D., & Crowell, A. (in press). What are the cognitive skills adolescents need for life in the 21st century? In J. Smetana & E. Amsel (Eds.), *Adolescence: Prospects and possibilities.* Taylor & Francis.

Kuhn, D., & Felton, M. (2000, January). *Developing appreciation of the relevance of evidence to argument*. Paper presented at the Winter Conference on Discourse. Text, and Cognition, Jackson Hole, WY.

Kuhn, D. & Franklin, S. (2006). The second decade: What develops (and how)? In D. Kuhn & R. Siegler (Eds.), (W. Damon & R. Lerner, Series eds.), *Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language.* (6th edition). Hoboken NJ: Wiley.

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing

argument skills in a computer-supported environment. *Child Development, 79 (5),* 1310-1328.

Kuhn, D., Pease, M., & Wirkala, C. (2009). Coordinating effects of multiple variables: A skill fundamental to causal and scientific reasoning. *Journal of Experimental Child Psychology, 103,* 268-284.

Kuhn, D., & Udell,W. (2003). The development of argument skills. *Child Development, 74*(5), 1245–1260.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning, 13*, 90–104.

Larson, A.A., Britt, M.A., & Kurby, C. (in press). Improving students' evaluation of informal arguments. *Journal of Experimental Education.*

Leitao, S. (2000). The potential of argument in knowledge building. *Human Development (43),* 332-360

Leitão, S. (2003). Evaluating and selecting counterarguments. *Written Communication, 20*, 269-306.

Moshman, D. (2005). *Adolescent psychological development: Rationality, morality; and identity* (2nd ed.). Mahwah NJ: Erlbaum.

Oaksford, M., Chater, N., & Hahn, U. (2008). Human reasoning and argumentation: The probabilistic approach. In J. Adler & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations.* New York: Cambridge University Press.

Sedlak, A., & Kurtz, S. (1981). A review of children's use of causal inference principles. *Child Development, 52*, 759-784.

Toulmin, S.E. (1958). *The uses of argument.* Cambridge: University Press.

Walton, D.N. (1989). Dialogue theory for critical thinking. *Argumentation, 3,*169-184.

Walton, D.N. (1996). *Argumentation schemes for presumptive reasoning.* Hillsdale, NJ: Lawrence Erlbaum.