

# Developing DiALoG: A Digital Formative Assessment Tool to Support Student and Teacher Learning of Oral Argumentation

April B. Holton, J. Bryan Henderson, Arizona State University  
Megan Goss, Eric Greenwald, Lawrence Hall of Science, UC Berkeley  
Barbara Barcus, Amy Lewis, Peoria Unified School District, Arizona  
april.holton@asu.edu, mgoss@berkeley.edu, BBarcus@pusd11.net, ALewis@pusd11.net,  
egreenwald@berkeley.edu, jbryanh@asu.edu

**Abstract:** We present lessons learned from an ongoing collaboration between researchers and practitioners to test and refine a teacher formative assessment tool named DiALoG (**D**iagnosing **A**rgumentation **L**evels of **G**roups). DiALoG is a tablet-based instrument that enables teachers to score oral classroom argumentation in real time across multiple dimensions. Coupled with the DiALoG tool are Responsive Mini Lessons (RMLs), which provide follow-up lessons for teachers to act on assessment information from the tool to increase proficiency with the dimensions of argumentation assessed. Lessons learned include evidence that teachers' use of DiALoG promotes attention to particular aspects of classroom interactions during argumentation sessions that the teachers had not previously considered, and along with the RMLs, may help fill gaps in their pedagogical content knowledge. Teacher feedback on the user interface has informed the design of future iterations of the formative assessment instrument.

## Introduction

Concern that the shallow memorization of science facts should not be the singular learning goal in the science classroom has caused education researchers and educators around the world to call for a shift in how students learn in the science classroom (National Research Council, 2011). In short, this shift requires that students engage actively in learning science by investigating phenomena to develop and defend evidence-based explanations of the natural world. In the United States, for example, the National Research Council's *Framework for K-12 Science Education* (2011) and the *Next Generation Science Standards* (NGSS Lead States, 2013) call for more *authentic* learning experiences, and that these experiences are better aligned to actual scientific practices. In Australia, researchers found that a concern for improving the Science, Technology, Engineering and Math (STEM) capacity of their students has caused many countries to make dramatic changes in the ways in which science is taught, including the implementation of science programs that place an emphasis on critical thinking and science practices (Marginson, et al., 2013). Furthermore, the European Union officially recommended incorporating scientific argumentation as a set of key competencies for lifelong learning (European Union, 2006). The Program for International Student Assessment (PISA) considers the ability to use scientific evidence to support claims – a skill that is fundamental for engaging in scientific argumentation – to be a key competency in their evaluation of students worldwide (OECD, 2006; 2013).

The practice of scientific argumentation supports a more evidence-centered and less memorization-focused classroom. Scientific argumentation is defined in the United States' NGSS as '...a process for reaching agreements about explanations' (Lead States, 2013, Appendix F, p.13). Successfully enacting the practice of argumentation involves facility with a wide range of communicative acts, including: "...conjectures, conclusions, explanations, models...scientists must also convince others that their evidence is relevant and of high quality so they spend lots of time assessing, critiquing, justifying and defending the evidence" (Sampson, Enderle, Grooms, 2013, p. 31). Argumentation can be enacted in the classroom through a variety of activities where students are asked to talk, read and/or write about claims and evidence. Many studies have found that it is both challenging to teach and for students to learn this rich and complex practice (Driver, Newton, & Osborne, 2000; Duschl, et al., 2007; Duschl & Osborne, 2002).

One way to support both teachers and students with enacting and improving argumentation in their classrooms is by providing a way to formatively assess specific aspects of this practice. The oral modality has proven particularly difficult to assess in real-time and traditionally receives less attention than reading and writing. Relatively few valid and reliable assessments for speaking and listening exist (Popham 1997, 2011), and next to none that focus on the construction and critique of scientific arguments (Ford, 2008). Our study addresses the use of educational technology to support teacher and student growth in oral argumentation. DiALoG (**D**iagnosing **A**rgumentation **L**evels of **G**roups) is a formative assessment instrument in the form of an iPad application for teacher use in real-time

observation of oral argumentation. And, to further support teachers, the scores from the DiALoG tool correspond with **Responsive Mini-Lessons (RMLs)** that teachers can use to address gaps in students' knowledge and abilities in order to increase their oral argumentation capacity.

The purpose of this practitioner-oriented paper is to share the design and components of the DiALoG formative assessment tool, the corresponding RMLs, and the insights gained through our collaboration with teachers. In addition, we will share our thinking about the implications for future design and study, especially as these implications relate to supporting teacher learning about the practice of argumentation in the classroom and how formative assessment can be used to support both student and teacher learning.

## **The DiALoG (Diagnosing Argumentation Levels of Groups) assessment and accompanying RMLs (Responsive Mini-Lessons)**

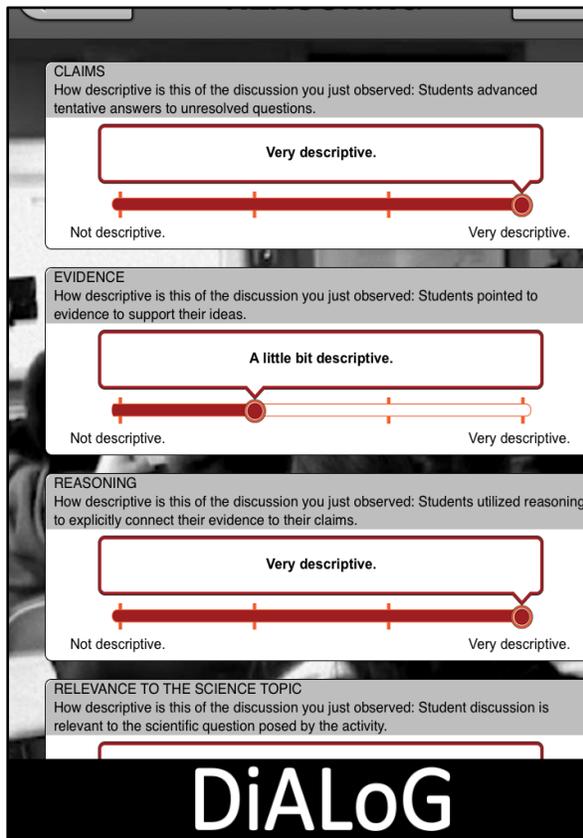
The DiALoG tool and the associated lessons are central components of a 4-year study, *Diagnosing the Argumentation Levels of Groups (DiALoG): A Digital Formative Assessment Tool to Support Oral Argumentation in Middle School Science Classrooms* that was funded through a United States National Science Foundation DRK-12 government grant (Award #1621496). The tool and follow-up RMLs were created to provide a practical way for teachers to formatively assess their students' oral argumentation abilities and, simultaneously, better understand the complex discourse themselves. To make scoring user-friendly, a tablet-based application was developed for teachers to enter scores flexibly with the touch of a finger as they move throughout the room listening to students as opposed to entering, deleting, and re-entering scores when they have a change of mind.

DiALoG assesses two bundles of items based on Erduran, Simon, and Osborne's (2004) distinction between *argument* (i.e., the argument product that results from the argumentative process) and *argumentation* (i.e., the argument process itself): intrapersonal arguments and interpersonal argumentation.

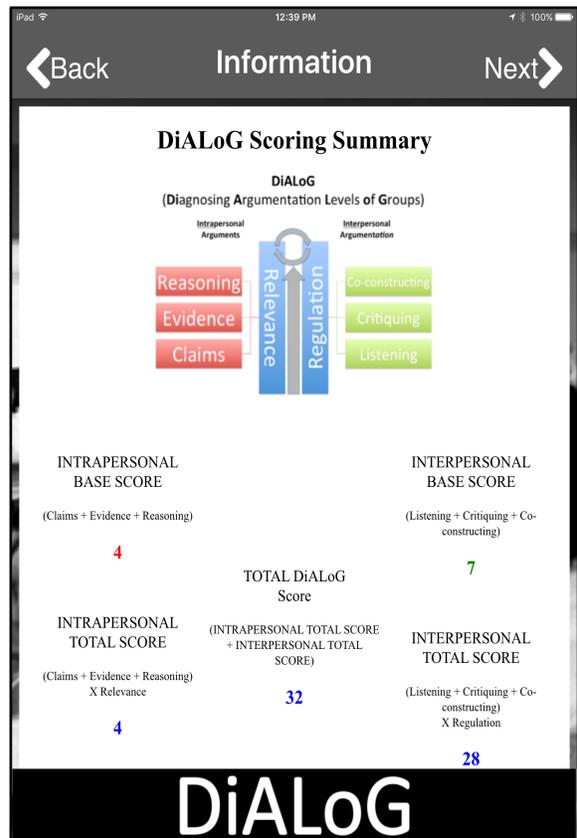
Intrapersonal arguments. The intrapersonal bundle of items is commonly known by science educators as the *Claims-Evidence-Reasoning*, or CER, framework (McNeill & Krajcik, 2011). When using the tool, teachers assign a score: 0 (Not Descriptive), 1 (Somewhat Descriptive), and 2 (Very Descriptive) on items assessing the degree students verbally offered tentative answers to scientific questions (i.e., *Claims*), justified claims by citing *Evidence*, and linked evidence to claims through *Reasoning* (see Figure 1). Because it is possible for students to articulate arguments that are scientifically unsound or irrelevant, DiALoG contains a fourth item gauging the *Relevance* to the science content of the lesson. This *Relevance* item is multiplied by the sum of the scores for the *Claims*, *Evidence*, and *Reasoning* items.

Interpersonal argumentation. The interpersonal component focuses on the quality of social interaction between students when they speak in groups. The *Listening* item probes the degree in which student contributions reflect genuine listening to what other speakers have previously articulated. The item probing *Critiquing* assesses the willingness of students to push the thinking of others through critical questioning. Meanwhile, the *Co-constructing* item measures the degree students collectively deepen their thinking by integrating their individual constructions with constructions of others. As with the Intrapersonal dimension, the Interpersonal dimension contains a fourth item multiplier, *Group Regulation*, which assesses the level of awareness participants have of the goals and participation of the group as a whole. The DiALoG application instantly calculates scores for both the Intrapersonal and Interpersonal dimensions (see Figure 2), and hence, has the potential to be a powerful feedback device for teachers to quickly identify the strengths and weaknesses of the classroom talk they observe.

Recognizing that formative assessment is most powerful when teachers take action on the information they gather, a portfolio of Responsive Mini Lessons (RMLs) aligned with different possible scores for each item were developed. DiALoG provides instructors a picture of the strengths and weaknesses of their students, as seen through the numbers assigned for each item. From here, a teacher can complete the next phase of the formative assessment loop, by choosing a targeted RML that provides a teaching and learning opportunity for students to become better at a specific item. These RMLs take about 30-60 minute to implement as a follow-up lesson. For example, if the teacher decides from the scores via the DiALoG tool that students are in need of more support with listening when engaged in oral discussions, they can choose to teach a RML that corresponds to the *Listening* dimension.



**Figure 1:** Screen shot of the DiALoG user interface. Teachers swipe their finger on the touchscreen to rate how descriptive each item is of what they are observing in their classroom. The ratings can easily be adjusted as they gather more evidence on the performance of the class as a whole.



**Figure 2:** Screen shot of the score breakdown for both the Intrapersonal and Interpersonal dimensions of DiALoG. This summary can provide teachers with a quick way to identify the relative strengths and weaknesses of different aspects of observed classroom discussions.

## Collaboration with teachers

Five middle school science educators whose classroom experience ranged from one to more than 20 years were recruited to collaborate with researchers by piloting and providing important feedback for the iterative refinement of DiALoG and the corresponding RMLs. Teachers participated in an initial professional development session familiarizing them with the DiALoG tool and its eight target items of classroom discourse. Teachers viewed classroom videos and practiced using DiALoG to rate the discussion and prescribe an appropriate RML. Following the professional development sessions, collaborating teachers agreed to allow our research team to conduct classroom observations, interviews, and surveys.

Researchers observed multiple classroom sessions for each teacher, including when teachers used DiALoG to assess oral argumentation during one of their own lessons, as well as when teachers implemented an RML that was chosen based on areas of improvement identified by DiALoG. Classroom observations focused on the frequency of DiALoG use, teacher behaviors, and observer's perceptions of classroom successes and challenges when using the tool. Following each classroom observation, researchers interviewed the collaborating teachers. Teachers shared their impressions about both the user interface of DiALoG and the targeted dialectical items. During these conversations, teachers made suggestions for improvements based on what they had just experienced with their students. In addition, they reflected on how the use of DiALoG impacted their own instruction. Furthermore, teachers explained their rationale for selecting specific RMLs based on information provided by DiALoG.

Pilot teachers also completed a 23-item survey designed to capture information on their comfort and practice with argumentation and formative assessment in general. The survey used Likert-type items asking teachers to rate the degree to which they agreed or disagreed with statements, such as "I am comfortable giving students control during

peer-to-peer discussions” or “Students learn more science when they argue with one another about the ideas they might have.” Additionally, teachers responded to open response items focused on their epistemological orientations about the nature of science and how they believed science content is best taught. Examples of these items included: “Providing opportunities for middle school students to engage in scientific argumentation is important because:” and “What do you think makes an argument scientific?”

Teacher recommendations for improvements to the DiALoG assessment, including the design of the software user interface and the design of the accompanying RMLs, were collected throughout the various conversations and encounters described above. This important teacher feedback influenced and continues to influence iterative revisions of DiALoG, RMLs, and supports provided for teachers to help them better understand the reflexive relationship between DiALoG and RMLs. For example, wireframes for a new design of the DiALoG user interface were created based on teacher feedback. These new designs were shared with collaborating teachers, who provided additional feedback prior to development of the new interface.

## Important insights from pilot work with teachers

As mentioned previously, teachers during pilot field-testing have provided valuable input on the iterative refinement of the DiALoG assessment and the accompanying RMLs. Furthermore, pilot work with teachers is beginning to suggest that when instructors pay closer attention to DiALoG items and implement appropriate RMLs to support growth, they too learn more about argumentation to improve their practice. Based on data gathered through observations, interviews, and surveys, highlighted below are examples of important teacher insights as they relate to opportunities and challenges for the DiALoG project.

## Opportunities

*Teachers are an integral part of refinement of DiALoG and the RMLs.* Our close collaboration with teachers revealed several opportunities for refinement. For example, during the initial professional development sessions, teachers called our attention to a possible confound with the listening dimension. Having viewed a video of classroom argumentation, teachers noted that, based on their following of the instrument prompts, scores for listening were lower than their instincts were telling them. The DiALoG criterion for high listening score included that students display “respectful” body language. After discussing this confound, it was concluded that when it comes to the student actions that demonstrate civility and respect, they are more of a matter of scoring the quality of DiALoG Interpersonal dimension, *Group Regulation*.

In addition, teacher suggestions drove the redesign of the DiALoG user interface. The initial design split the Intrapersonal and Interpersonal items on two different tablet screens. Multiple teachers mentioned that they would prefer to be able to score all eight items of the DiALoG assessment on a single screen. One teacher explained, “It would be helpful to be visually reminded of the existence of, and the relationship between, all of the dimensions, in order to take an integrative approach to my assessments.” Furthermore, multiple teachers expressed a desire to be able to open and save multiple DiALoG sessions at once to use in small group sessions. Finally, teachers voiced the importance of annotating as they made observations using the tool. They felt this would be helpful for reflection to choose an appropriate RML as a follow up.

*The DiALoG instrument helps teachers shift to a student-centered mindset.* Survey results indicated that none of the teachers felt “very comfortable” with “giving students control during peer-to-peer discussions.” Multiple teachers mentioned in interviews that, prior to using DiALoG, their tendency during student discussion was to look for specific student responses and/or anticipate what they would say to students next. When scoring using DiALoG, the instrument prompted teachers to listen more closely to the totality of what students were saying and restrain from interjecting into student conversations as readily. When teacher interjections were made while using DiALoG, they were more often based on direct feedback from DiALoG scores.

*The DiALoG instrument draws teachers’ attention to specific gaps in their own understanding of student discourse.* Survey results indicated that teachers valued opportunities for students to engage in scientific argumentation. However, the data also suggested that teacher implementation of argumentation was at a relatively novice level prior to being introduced to DiALoG and the accompanying RMLs. Despite this inexperience, DiALoG supported teacher growth by promoting more nuanced awareness of the multiple important aspects of successful classroom argumentation. When asked during interviews if and how DiALoG was useful, teachers acknowledged that it helped focus their attention to specific aspects of students’ argumentation. Without the tool, they tended to assess classroom discussion by how “well” it was going in a general sense. Multiple teachers noted that without DiALoG’s prompting, they would not have thought to assess the quality of student critique or co-construction of ideas.

## Challenges

*There is variation in how teachers interpret DiALoG scores.* DiALoG was developed as a formative assessment and guide for selection of follow-up RMLs, and multiple teachers used it in this fashion. However, a few pilot teachers wanted to use the assessment in a summative fashion. In some cases, teachers were concerned with how assessment at the group level translates to an individual grade. This concern with grading presents a challenge, as the assignment of specific scores can negate the purpose of formative assessment. More specifically, the essence of formative assessment is to utilize real-time information to move learning *forward*, but a specific grade suggests looking *backward* to assess what has already transpired. Indeed, formative assessment research (Butler, 1988; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006) suggests that when feedback is provided as a number or grade, students do not internalize or take action on improving performance moving forward. In contrast, when constructive feedback or suggestions for improvement are provided instead of a mere score or grade, students are more likely to take future action. We are finding a similar trend for teachers when they focus on numerical scores for each DiALoG item. Indeed, the numbers generated by DiALoG are leading some teachers to use the instrument as a summative tool, as opposed to a formative assessment. Because DiALoG was designed to formatively assess oral argumentation at the group level, this feedback helped us to see that we need to provide more rationale about why the tool is most useful when applied to a group discussion. Furthermore, summative grading of individual students has the potential to draw teacher attention away from more important general trends in student discourse.

*Teachers could benefit from metacognitive reflection as they develop proficiency with DiALoG.* DiALoG and the accompanying RMLs were developed for teachers to use without necessary intervention in the form of professional development or instructional coaching. Several teachers were notably less concerned with connecting DiALoG scores to lessons or teaching strategies. Rather, they envisioned using DiALoG primarily as a grading tool. These instructors demonstrated perhaps a less sophisticated grasp of formative assessment than other pilot teachers, which emphasizes the value of providing explicit encouragement to reflect on and connect assessment ratings to pedagogical strategies. Without opportunities for professional development, this suggests a possible design enhancement that structures explicit reflection to connect DiALoG scores to the RMLs, so that the formative assessment-to-action cycle is more structurally supported.

## Implications for future study

Several valuable insights guided implications for our future study. First, ease of use is a primary focus, so interface changes were made based on important feedback from pilot teachers. The next iteration of DiALoG will include the ability for teachers to toggle between items on a single screen and to enter annotations as they score. This will make it easier for teachers to make connections between the Intrapersonal and Interpersonal dimensions of oral argumentation, as well as more accurately reflect on what they have observed.

While a user-friendly software interface is important, we want to ensure that depth is retained to allow for thoughtful use, and finding the right balance has implications for continued design. While the interface provides a quick way for teachers to ascertain the levels of each item and dimension in the form of scores, as well as an overall score (see figure 2), this has consequences in how the teacher uses the assessment. Based on the previously mentioned research suggesting that students are more likely to improve when provided specific feedback rather than a grade or number and our experiences with teachers' emphasis on assigning a grade, the next version of DiALoG item scoring levels will be represented by statements rather than scores ranging from 0-2. Teachers will score each item as "Blank stares", "There, but rare", and "There, everywhere." Data will be gathered to determine if using the words instead of scores makes a difference with teachers' tendency to use the information provided for summative evaluation, rather than formative. Furthermore, there is a concern that the focus on grades or assigning a numerical score as a summative assessment may distract teachers from the ultimate goal of building a discourse community to support scientific thinking.

In addition, the desire for teachers to track individual students to meet their grading goals needs further consideration. While the interface can be designed to toggle between groups, pairs, and whole group, the assessment was designed to assess the whole group. We would like to continue to consider if assessing beyond the whole group is beneficial.

Since many future users of DiALoG most likely will not receive extensive professional development on assessing oral argumentation, future support materials may need to be more explicit about the difference between formative and summative assessments, and the explicit purpose of DiALoG and the accompanying RMLs. Reflection is the bridge between the observations made using DiALoG and the choice of RML to move student and teacher learning forward. Therefore, we will study how the annotating feature and user guides can support teacher reflection.

In conclusion, the DiALoG formative assessment system provides concrete, actionable insight into learning for both students and for teacher practice. Additional research is necessary to determine to what degree the use of DiALoG and the corresponding RMLs helps shift teachers' mindset about argumentation and formative assessment. Investigation of this phenomena will continue on a larger scale in the following years of the grant.

## References

- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task- involving and ego- involving evaluation on interest and performance. *British journal of educational psychology*, 58(1), 1-14.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, 84(3), 287-312.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. E. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
- Duschl, R.A. and Osborne, J. (2002) Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, 88(6), 915-933. <https://doi.org/10.1002/sci.20012>
- European Union (2006). Recommendation of the European Parliament on key competences for lifelong learning. *Official Journal of the European Union*, 30-12-2006, L 394/10-L 394/18.
- Ford, M. J. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, 92, 404-423.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Marginson, S., Tytler, R., Freeman, B., and Roberts, K. (2013). STEM: country comparisons: international comparisons of science, technology, engineering and mathematics (STEM) education. Final report. Australian Council of Learned Academies, Melbourne, Vic. Retrieved from <http://dro.deakin.edu.au/eserv/DU:30059041/tytler-stemcountry-2013.pdf>
- McNeill, K.L., & Krajcik, J.S. (2011). *Supporting grade 5-8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing*. Pearson.
- National Research Council. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, D.C.: The National Academies Press.
- NGSS Lead States. (2013). Next generation science standards: For states, by states. Retrieved from <http://www.nextgenscience.org/>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self- regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.
- Organisation for Economic Cooperation and Development (OECD) (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
- Organisation for Economic Cooperation and Development (OECD) (2013). *PISA 2015 Draft Science Framework*. Paris: OECD.
- Popham, W.J. (1997). What's wrong and what's right with rubrics. *Educational Leadership*, 55, 72-75.
- Popham, W.J. (2011). *Transformative assessment in action: An inside look at applying the process*. Association for Supervision and Curriculum Development (ASCD) Alexandria, VA.
- Sampson, Victor, Enderle, Patrick, & Grooms, Jonathon. (2013). Argumentation in Science Education. *Science Teacher*, 80(5), 30-33.