Towards argumentative grammars of design research

Arthur Bakker

Bremen University, Germany; Utrecht University, the Netherlands; a.bakker4@uu.nl

Design research is considered a valuable but demanding methodological framework that continues to generate theoretical and methodological reflection. An important topic to address is that of argumentative grammar, the logic guiding a method and supporting warranted claims, because critics consider it a weak spot of design research. With reference to the history of logic, I challenge these critics' demand of an argumentative grammar that relies solely on structure rather than also on content. The purpose of this paper is to think through what argumentative grammars of design research could look like. Because the literature is so limited on this topic, I draw on interviews with experts in design research to evaluate and discuss my own attempt to formulate an argumentative grammar in relation to possible research questions. One conclusion is that design research requires multiple argumentative grammars depending on the design and the research focus.

Keywords: Design-based research, expert interview study, methodology.

The need for an argumentative grammar of design research

In a special issue on design research, Kelly (2004) argues that design research (DR) is a valuable emerging set of methods in education, but he has methodological concerns. In his view, "[t]he next task is to establish the logos of design research so that we can argue, methodologically, for the scientific warrants for its claims." (p. 105) For design research to become a methodology (method + logos), he proposes, we need an "argumentative grammar," which he defines as "the logic that guides the use of a method and that supports reasoning about its data" (p. 118).

A methodology that already has a clear argumentative grammar is that of randomized field trials as introduced by Sir Ronald Fisher in the 1920s for agriculture. In such trials, also called randomized controlled trials, researchers randomly attribute objects or subjects to an experimental or control condition so that they can assume that these two groups are equal on average except for receiving the treatment or not. Any differences between these two groups as measured by means of pre- and posttests can therefore be attributed to the difference in treatment. One advantage of this methodology is that its argumentative grammar is a structure that can be

> described *separately* from its instantiation in any given study so that the logic of a proposed study and its later claims can be criticized. Thus, many reviewers reject studies not on the choice of method (procedure), but on their violation of the underlying logos that one expects to see with that choice of method. (Kelly, 2004, p. 118, emphasis in the original).

Kelly and other critical friends such as Shavelson et al. (2003) thus push design researchers to make warranted claims and go beyond purely narrative accounts. With Cobb et al. (2014), I think design researchers should indeed work towards an argumentative grammar (or grammars) to increase DR's methodological quality. However, drawing on the history of logic, I problematize the preference for a *separate* logical structure for design research (DR) that is irrespective of content. The purpose of this paper is then to think through what an argumentative grammar for design research could look like instead. Because the literature is so scant on this topic, I decided to interview experts in DR on a provisional grammar that I formulated myself.

Problematizing the need of a separate argumentative structure

This section first addresses randomized field trials (RFTs) as the best known example of a methodology with an argumentative grammar that relies on the structure of argumentation. Next I use the history of logic to argue that logic that is based only on the structure and not on its content is of limited scientific value. I argue that DR requires argumentative grammars that acknowledge content as part of their logic, where content can refer to many things including key concepts used in the research, information on local circumstances (context), and the content of what is learned or aimed to achieve.

Randomized field trial

With respect to RFTs, I highlight three themes to prepare the discussion of an argumentative grammar of DR. Theme 1: In an RFT the design of an intervention and the evaluative research are separate. From some theoretical perspective or hypothesis, educational design with particular characteristics is developed and then evaluated. When the effects of its implementation are positive, these are attributed to characteristics of the design, operationalized in particular variables. Where does this leave the design researcher who typically intertwines intervention and evaluation?

Theme 2: Typical research questions reveal the type of knowledge that RFT are often after: What are the effects of intervention I on D? Is intervention I1 better than I2? The design researcher typically asks how particular learning can be supported or how some problem can be resolved (of course RFTs can also focus on mechanisms rather than just effects).

Theme 3: RFTs have a clear argumentative structure that is separate from the content of what is researched. This allows the audience, even those who may not have the expertise to engage with the content of the studies (e.g., key concepts, learning content, situation, mechanisms) to judge the structure of the procedure and the scientific reasoning. What could an argumentative grammar of DR look like if it, as I argue later, cannot depend on structure alone?

Despite their power, RFTs also have their limitations (MRC, 2000). To know "what works" in general is of little value if it is unknown "how and under what conditions things work" or what the mechanisms or "active ingredients" are that make an intervention work (Biesta, 2007). It is further acknowledged that valid measurement is difficult and that RFTs typically have good internal but not necessarily good external validity (cf. Shavelson, 2008).

Logic: Content matters as much as structure

I now use examples from logic to clarify that Kelly's (2004) and others' focus on the *structure* of argumentation may hold back educational research. The discipline of logic started in Aristotle's *Prior Analytics* with syllogisms such as "All men are mortal; Socrates is a man; therefore Socrates is mortal." This logic purely depends on the structure of the inference: The non-logical terms such as "mortal," "Socrates," and "man" can all be replaced by other terms without loss of validity. The interpreter does not even need to know the meaning of these terms to judge the validity of the reasoning. This reasoning is thus rigorous but irrelevant in scientific reasoning:

> This kind of logic based on syllogisms came into disrepute in the seventeenth century when science was born. Scientists like Descartes found that all interesting propositions, all interesting inferences are in fact nonsyllogistic. (Lakatos, 1999, p. 39)

Logic has developed in multiple ways. One nonsyllogistic type of reasoning relevant to science is what Brandom calls non-monotonic. This means that new conditions can turn a valid inference into an invalid one. Brandom (2000, p. 88) gives an example from physics:

1. If I strike this dry, well-made match, then it will ignite. ($p \rightarrow q$)
2. If p and the match is inside a very strong electromagnetic field, then it will *not* ignite." ($p \& r \rightarrow \neg q$)
3. If p and r, but the match is in a Faraday cage, then it will light. ($p \& r \& s \rightarrow q$)
4. If p and r and s and the room is evacuated of oxygen, then it will not light. ($p \& r \& s \& t \rightarrow \neg q$)

Scientific reasoning in educational research is clearly non-monotonic: There are overwhelming numbers of factors that can influence learning. Any relevant positive factor can probably be counteracted by a negative one. Given the pragmatic nature of education, it is also worth mentioning progress on pragmatic reasoning: Walton et al. (2008), for example, identified 96 argumentation schemes that people use in reasoning. It has also become evident that valid argumentation does not depend purely on structure but also on content (and context). So-called material inferences even purely depend on content rather than on their structure. Brandom (2000, p. 85) uses the inference from “Pittsburgh is to the west of Philadelphia” to “Philadelphia is to the east of Pittsburgh,” as an example of an inference that is materially valid because it depends only on the content of the concepts of east and west.

These brief observations from logic suggest that scientific progress relies not only on the structure of argumentation but also on content. Why then should research methodology in education be judged by the separate structure of its argumentation? But what would an alternative look like? Design researchers are faced with the challenge to come up with an alternative grammar or, more likely, grammars. One attempt is that by Cobb et al. (2014):

1. Demonstrating that the students would not have developed particular forms of mathematical reasoning but for their participation in the design study.
2. Documenting how each successive form of reasoning emerged as a reorganization of prior forms of reasoning.
3. Identifying the specific aspects of the classroom learning environment that were necessary rather than contingent in supporting the emergence of these successive forms of reasoning. (p. 490)

The function of such grammars is that they “link research questions to data, data to analysis, and analysis to final claims and assertions” (p. 489). Given that little has been written about this, I formulated an argumentative structure myself based on discussions with Karel Stokking and my own experience with doing and supervising DR. The most efficient and sensible way to gauge its quality seemed to be an interview study with expert design researchers. In this way I could explore what they thought about the need for an argumentative grammar of DR, what they thought of my attempt, and how it could be improved.

Method: Interview study with experts

I interviewed eighteen well-known international design researchers on argumentative grammars of DR and related themes for about 60-90 minutes. These experts represent a variety of different

disciplines and traditions in DR (seven were mathematics educators). Before presenting my own grammar proposal, I asked them about issues that might elicit their view on the logic accompanying DR and the type of claims it renders. First, I asked about the intertwinement of design and research because it can make particular claims difficult: In line with the argumentative structure of RFTs and thinking in terms of variables, many researchers prefer to keep design of an intervention and the (evaluative) research separate. Second, I asked experts' views on Kelly's (2004) claim that DR has no clear argumentative grammar. Third, I asked if they had a preference for types of research questions (what- vs. how-questions). An example of a what-question I showed to the interviewees is: "What are characteristics of a valid and effective teaching and learning strategy to teach students about correlation and regression in such a way that they experience coherence between mathematics and the natural sciences?" (Dierdorp, 2013). A how-question I presented is: "How can students be fostered in their connecting of gene as a molecular-level concept to phenomena at higher levels of biological organization?" (reformulation of Van Mil's, 2013, question). Fourth, I asked experts about the argumentative grammar I propose in the next paragraphs.

The focus on *how* to support learning in DR implies that in my view at least four things need to be captured in an argumentative grammar of a DR project. First, learning goals need to be underpinned (or a problem or needs analysis should be done). A design criterion could be relevance and a research criterion content validity (Plomp & Nieveen, 2013). Several existing methods (review study, expert interviews, Delphi study) can be used to this end. Second, a design (e.g., tool, teaching-learning strategy, or program) could be described in relation to theoretical and empirical considerations. Criteria here can be "empirically and theoretically underpinned" and "innovativeness," but some may want to emphasize "feasibility or practicability." Third, only if intentions are realized, particular intended phenomena can be studied (e.g., whole-class scaffolding; Bakker & Smit, 2017). In RFTs, the criterion would be formulated as "implementation fidelity," necessary to check if any effects can be attributed to the intervention having particular characteristics (cf. Sandoval, 2014). Fourth, information about to what extent learning goals are achieved, or a problem solved, needs to be given in order to answer the main question. The main criterion here is effectiveness.

The structure of a DR project presented to all interviewees for their feedback was the following:

> *How can goal X be achieved for a particular group of learners* (in particular conditions or under particular constraints)? To answer this main question, a sensible list of research questions could be:
>
> 1. What is an appropriate learning goal for….?
> 2. What is a design that would help students/teachers to achieve this goal?
> 3. How well was this strategy/trajectory implemented?
> 4. What were the effects of this intervention?

In discussing this structure, several topics arose that are related to aspects of argumentative grammar such as links between different parts of research (data, claims), in particular in contrast to RFTs. I summarize the experts' responses in three themes.

Theme 1: Intermingling design and research

A key feature of DR is that design and research progress hand in hand. In response to this issue, the interviewees noted the following points. First, any natural scientist knows that scientific practice, in particular the context of discovery, is much messier than presented in textbooks or reports of experiments. Of course, there is a place for experiments, but a large part of science—even in physics—is trial and error with set-ups, designing new arrangements, philosophizing, thought experiments et cetera. In certain disciplines, take astronomy, experiments are even impossible. Serendipity (e.g., the discovery of penicillin) also points to the importance of the context of discovery. The relative importance of RFT as a methodology rests on the side of justification. Several experts said that RFT-type research often produces "false security" or that it struggles with similar issues as other types of research, but somehow it has become common practice to ignore particular problems or trust researchers on doing it well (e.g., validity of measurement, identifying relevant variables). However, many noted there is also a place for RFT as it helps for example policymakers to decide between various well-established options to be implemented.

Second, two interviewees emphasized that DR is about how education *could* be. Where much research is about current educational practice, and some about its past, DR is about its future. Design researchers may argue that educational goals should be different from current educational practice, and design for these new goals. Such DR is thus after proofs of principle, not proof of doing better than current practice which may have very different goals. Comparison with a control group that worked towards different learning goals would be unfair. The argumentative grammar of this type of "proof of principle" DR is thus clearly different from DR that aims for causal claims about effectiveness of particular means of support. This points to the need for multiple grammars.

Third, several interviewees noted that DR conceptualizes learning environments as ecologies rather than systems that can be captured with a few manipulable and unmanipulable variables. Attributing an effect to particular variables then becomes challenging. Rather the focus should in the experts' view be on design principles, hypothetical learning trajectories, or mechanisms of learning, in line with DR's intention to produce knowledge about *how* things work (cf. Sandoval, 2014).

Theme 2: Research questions

Most interviewees considered the examples of what- and how-questions presented to them as too broad. Some did not have a strong preference for either formulation: The researcher wants to know similar things in both cases. However, most experts preferred the how-questions because these emphasize the process of achieving particular learning goals or solving a particular problem. In terms of Cobb et al. (2003), DR typically aims to provide insight into how particular means can support particular learning. This hints at the type of knowledge claims that DR purports to deliver.

A view, expressed by Abrahamson and diSessa, was that DR is a methodological framework (not a method or a strategy) that provides a generative context (about how education could be). Because new types of learning are promoted, new phenomena may emerge and thus in turn become objects of investigation. This view fits with the image of DR as a context of discovery for researchers. Once such phenomena are implicated and objectified, they can be studied as interesting in their own right, with little or no reference to the broader design research context (e.g., Abrahamson et al., 2016). In

line with the generativity of DR, many interviewees emphasized that interesting research questions often emerge rather late in the research process. They are hard to formulate in advance.

Theme 3: Argumentative grammar

The interviewees were overall positive about the proposed grammar. The elements of learning/educational goals, design, implementation, and effects are key to DR, and can be studied empirically, perhaps even in separate publications. One interviewee expressed some resistance to categorizations and structures in research because each project is unique and requires flexibility and creativity. Yet structures could be useful to early career researchers as a starting point.

The experts' further comments were matters of detail. With regard to the learning goals, diSessa noted that he sometimes preferred learning goals that colleagues thought were impossible to achieve with certain age groups (e.g., comprehending velocity and acceleration as vectors in Grade 6). McKenney pointed out that design researchers often encounter obstacles that can become the topic of research. She tends to do a lot of "front-end" work in the early phases of DR in areas where too little is known to arrive at effective designs.

Judging the quality of implementation was considered a good idea, although several experts noted that the implementation process could be interesting to study even without judging its quality. diSessa remarked that failure can be interesting from a design perspective. In his experience, many colleagues respond with surprise when he reports failure, but as long as important lessons can be learned, contributions to the knowledge base can be made.

The terms "interventions" and "effects" elicited some resistance due to connotations with the RFT paradigm of thinking in terms of variables. Several experts preferred to talk in terms of learning ecologies instead. However, some found it important to measure what was achieved and thought that design researchers had measured too little in the past. Many noted that there is certainly a place for RFTs, as well as for quantitative measurement, in DR. Some indicated that RFT ideally gives insight into mechanism too, and can be part of DR.

Ruthven suggested a fifth element, namely an improved re-design, which is indeed in line with DR's emphasis on the hypothetical status of any claims. Citing Cronbach, Plomp emphasized this holds for any type of research: "When we give proper weight to local conditions, any generalization is a working hypothesis, not a conclusion" (Cronbach, 1975, p. 125).

Plomp noted that although he did not write about argumentative grammar, his approach with Nieveen (Plomp & Nieveen, 2013) has such a function. For each phase of a DR project one criterion was central: relevance for the exploratory phase (e.g., problem analysis), consistency (of the design), practicability (of using the design), and last effectiveness.

An issue raised was whether different criteria were needed for DR than for some other research approaches. Because many readers and reviewers are used to different commissive spaces, experts such as Cobb stressed that DR has to become clear on the criteria on which it wants to be judged. For example, we have to acknowledge that design researchers are part of the research, and that their qualities as designers and researchers matter. As Confrey noted in the interview: "You build a reputation for doing good work (…), but that's not great for newcomers because they don't have the track record yet." It certainly goes against the more conventional norm of reliability that research

should be independent of the researcher. Hence it seems necessary to think through the criteria by which design researchers want to be judged. However, McKenney preferred the research part of DR to be treated with the same criteria as other qualitative or mixed-methods approaches. Kelly suggested DR can learn from other research approaches such as single-subject and repeated-measures designs.

Not only the design researcher, but the audience has to make judgments as well. Where RFTs can yield results that sometimes seem to require little understanding of the topic at hand, DR asks for an audience that can appreciate the relevance of the educational goals chosen, the innovativeness of the design, and the learning processes reported. diSessa noted that the typical reasoning in DR is to show what types of reasoning can be promoted in a particular way, for instance by using particular software. Any well-informed domain-specific educational researcher with knowledge of the disciplinary (e.g., mathematical) content will know how rare or relevant such types of reasoning are for particular age groups, so will appreciate qualitative examples of even small samples.

Conclusion

In this paper I have argued that it is unreasonable to expect that educational research including DR should use an argumentative grammar that depends solely on structure rather than also content (key concepts, mathematical learning content, context etc.). Examples from logic illustrate the importance of types of reasoning that are also based on content. Argumentative grammars for DR should thus acknowledge content too. Cobb et al. (2014) offered an argumentative structure that can help convince readers about the development of students' mathematical reasoning and the aspects of the learning environment that supported them (see also Sandoval, 2014). My own proposal focused on the grammar of a DR project with the aim to contribute to knowledge about how particular educational goals could be achieved in general (or problems solved). Based on the interviews with experts, my proposal—after some modification—seems to make sense as a starting point for design researchers when they write a proposal or want to demarcate phases in their overall project (cf. McKenney & Reeves, 2012; Plomp & Nieveen, 2013) with criteria that are central in each phase. However, there is a need for more explicit argumentative grammars, for instance for "proof of principle"-type DR and for smaller-scale design studies that focus on interesting phenomena that are discovered during a larger DR project.

References

Abrahamson, D., Shayan, S., Bakker, A., & van der Schaaf, M. (2016). Eye-tracking Piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Human Development*, *58*(4-5), 218–244.

Bakker, A., & Smit, J. (2017). Theory development in design-based research: An example about scaffolding mathematical language. In: S. Doff & R. Komoss (Eds.), *How does change happen?* (pp. 109-124).Wiesbaden: Springer

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Educational Theory*, *57*(1), 1–22.

Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge, MA: Harvard University Press.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cobb, P., Jackson, K., & Dunlap, C. (2014). Design research: An analysis and critique. *Handbook of international research in mathematics education*, 481–503.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116–27.

Dierdorp, A. (2013). *Learning correlation and regression within authentic contexts*. Utrecht: FIsme Library.

Kelly, A. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences, 13*(1), 115–128.

Lakatos, I., Feyerabend, P., & Motterlini, M. (1999). *For and against method: including Lakatos's lectures on scientific method and the Lakatos-Feyerabend correspondence*. University of Chicago Press.

McKenney, S., & Reeves, T. (2012). *Conducting educational design research*. London: Routledge.

MRC (2000). *A framework for development and evaluation of RCTs for complex interventions to improve health*. London, UK: Medical Research Council.

Plomp, T., & Nieveen, N. (2013). *Educational design research: Illustrative cases.* Enschede: SLO.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36.

Shavelson, R. J. (2008) Guest editor's introduction. *Applied Measurement in Education, 21*, 293–294.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25–28.

Van Mil, M. H. W. (2013). *Learning and teaching the molecular basis of life*. Utrecht: FIsme Library.

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes. Fundamentals of critical argumentation.* New York: Cambridge University Press.