

AMEE GUIDE

Building bridges between theory and practice in medical education using a design-based research approach: AMEE Guide No. 60

DIANA H. J. M. DOLMANS & D. TIGELAAR

Maastricht University, The Netherlands

Abstract

Medical education research has grown enormously over the past 20 years, but it does not sufficiently make use of theories, according to influential leaders and researchers in this field. In this AMEE Guide, it is argued that design-based research (DBR) studies should be conducted much more in medical education design research because these studies both advance the testing and refinement of theories and advance educational practice. In this Guide, the essential characteristics of DBR as well as how DBR differs from other approach such as formative evaluation are explained. It is also explained what the pitfalls and challenges of DBR are. The main challenges deal with how to insure that DBR studies reveal findings that are of a broader relevance than the local situation and how to insure that DBR contributes toward theory testing and refinement. An example of a series of DBR studies on the design of a teaching portfolio in higher education that is aimed at stimulating a teacher's professional development is described, to illustrate how DBR studies actually work in practice. Finally, it is argued that DBR-studies could play an important role in the advancement of theory and practice in the two broad domains of designing or redesigning work-based learning environments and assessment programs.

Introduction

What is the problem?

Medical education research has grown enormously over the past 20 years worldwide. The number of scientific journals has increased and also the number of participants at international conferences on medical education research (Eva 2009). But, does medical education research lead to improvements in educational practice and does it contribute toward the advancement of knowledge?

Educational practitioners often complain that there is a gap between educational research and educational practice (Badley 2003). They argue that the research that is conducted within the general education domain is not relevant for educational practice. Within medical education research, it is often argued that medical education research does not contribute toward an increase of a body of knowledge or theory building (Albert et al. 2007). According to influential leaders in the field, a lot of studies are being reported in the medical educational journals that have already been done before and that do not add new knowledge. A number of studies lack a theoretical background (Albert et al. 2007; Norman 2007). This is a big problem. If studies are being conducted that do not rest on theories, then it will become very difficult to understand or explain the underlying factors or causes or to explain why an intervention works or does not work.

Practice points

- Good DBR does contribute toward both testing and refinement of theories and improving educational practice.
- Good DBR is characterized by a close interaction between practitioners, designers, and researchers.
- DBR uses a mixed-methods approach to understand underlying the processes or factors.
- DBR leads to design guidelines that specify which characteristics are crucial for a particular intervention in a specific context.
- DBR is a fruitful approach for design research especially when (re-)designing work-based learning environments and assessment programs.

Educational research, both in general education and medical education, should both contribute toward the building of a body of knowledge and theories (Reeves et al. 2008; Monrouxe & Rees 2009) and understanding problems encountered in educational practice (Eva 2009). Education takes place in a complex environment in which many variables interact with each other (Berliner 2002; Kember 2003). Education research should therefore not only be aimed at proving whether an intervention works or not but also should be aimed at understanding this complexity (Regehr 2010). Education research should be aimed at investigating how

Correspondence: D. H. J. M. Dolmans, Department of Educational Development & Research, Faculty of Health, Medicine and Life Sciences, School of Health Professions Education (SHE), Maastricht University, P.O. Box 616, Maastricht, The Netherlands. Tel: 31 43 388 5730; fax: 31 43 388 5779; email: d.dolmans@maastrichtuniversity.nl

different variables influence each other and why and how interventions work or do not work (Cook et al. 2008). Much more studies are needed in medical education research that focus on clarifying processes underlying the observed effects and answer the question “Why or how did it work?” (Cook et al. 2008).

In this Guide, it will be argued that design-based research (DBR) can help to bridge the gap between research and practice, because it does contribute toward theory testing and refinement on the one hand and improvement of educational practice on the other hand.

The theory behind DBR explained

What are the characteristics of DBR?

DBR is aimed at understanding the complex world of educational practice. One first important characteristic is that DBR does not take place in a laboratory setting, but in a real-life setting where learning normally takes place (Barab & Squire 2004). DBR studies’ focus on complex problems about the nature of learning in a real learning context (Collins et al. 2004). Another important characteristic of DBR is that it takes place in continuous cycles of design, evaluation, and redesign (Design-Based Research Collective 2003). Thus, a design is tested or evaluated in an authentic learning environment. The evaluation leads to new insights, based on which the design can be further improved or redesigned. A third important characteristic of DBR is that the design of the learning environment is based on theoretical principles. DBR is aimed at advancing theories (Barab & Squire 2004). Theories are very important in education research because they help us to understand how or why something works or does not work under particular circumstances in a complex environment such as education, in which many different variables interact with each other. In other words, good DBR advances theories and practice (Barab & Squire 2004). A fourth important characteristic of DBR is that not a single method is used to evaluate the design, but DBR is characterized by mixed-methods studies. In order to better understand the complex interactions between students, teachers, learning materials, and assessment tools, DBR makes use of mixed methods, both quantitative and qualitative methods that clarify what works under which circumstances. Mixed methods within one study are very useful when studying complex interactions in natural settings, because the findings may complement each other, make it possible to cross-check findings and make it possible to discover inconsistencies (Schiffedercker & Reed 2009). A fifth important characteristic of DBR is that designers, researchers, and practitioners with different expertise work closely together in designing, evaluating, and redesigning the learning environment (Barab & Squire 2004). They interact and communicate with each other frequently, share their ideas, and together refine theories and practice. So, designers, teachers, and researchers collaborate in conducting the research and disseminating it. This contributes toward creating ownership and commitment from teachers and learners. It also makes us clear on how the design is reinterpreted and adapted in

Table 1. Five important characteristics of DBR.

- Takes place in continuous cycles of design, evaluation, and redesign
- Takes place in authentic real-life learning settings where learning takes place normally
- Is aimed both at testing and refining theories and also advancing practice
- Is characterized by mixed-methods studies
- Designers, researchers, and practitioners with different expertise interact frequently and share their ideas

Source: Based on Barab and Squire (2004) and Design-Based Research Collective (2003).

educational practice (Design-Based Research Collective 2003). Table 1 summarized these characteristics.

A short history of DBR

DBR or design research was first described in the early 1990s (Reeves et al. 2005). In 1992, the term design experiments was introduced by Brown (1992) because she was of the opinion that design experiments could fill a niche in the methodologies that were used since that time to study learning (Collins et al. 2004). Design experiments were aimed at addressing theoretical questions about the nature of learning in a real learning context and at conducting formative evaluations that test and refine educational designs based on theoretical principles (Collins et al. 2004). Since that time, different terms are being used in the literature such as design experiments, design research, development research, and DBR (Design-Based Research Collective 2003; Barab & Squire 2004; Collins et al. 2004; Kelly 2004; Reeves et al. 2005; Van den Akker et al. 2006). The terms can be used interchangeably, but in this article, we use the term design-based research or DBR.

DBR and other approaches

After having read this Guide so far, some readers might come to the conclusion that DBR is a new word for formative evaluation or action research; i.e., research in which instructors study or evaluate their own teaching practice in order to improve it. DBR is different from formative evaluation studies or action research, even although these approaches might have a lot in common, such as the iterative and cyclic process of design, evaluation, and redesign. DBR is not only a matter of formatively evaluating whether a particular intervention meets certain basic standards or which aspects of the intervention need improvement. Good DBR is also aimed at testing or refining theories about the nature of complex learning environments (Design-Based Research Collective 2003). In other words, the interventions are closely connected to theories and the refinement and testing of theories is intertwined with designing and improving the learning environment. The connection with theory is the basic difference between DBR and formative evaluations or action research. DBR is aimed at advancing theoretical knowledge about learning and learning environments (Barab & Squire 2004). Theoretical foundations and claims are at the heart of the approach (Reeves et al. 2005).

DBR and mixed-methods research

What is also important to keep in mind is that DBR uses multiple methods; methods that provide insight in how different variables influence each other; and methods in which not only attention is paid to outcomes but also to the underlying processes that could explain why and how an intervention does work or not. Experimental studies in laboratory settings are not conducted within DBR. The problem is that controlled experimentation in educational interventions may lead to a rather reductionist and trivial exercise (Gruppen 2008). We do not argue that controlled experiments should never be done. The best method to be chosen is of course dependent on the research question formulated; in some cases, e.g., when testing the influence of one specific variable on learning, it might be a very useful method. But, if we want to answer research questions about the nature of learning in authentic learning environments that both enhance the testing and refinement of theories and educational practice, we should use a DBR approach and a mixed-methods approach. Mixed-methods research is an approach to research in which quantitative and qualitative research techniques, methods, or approaches are used in a single study (Johnson & Onwuegbuzie 2004). The idea is that researchers should collect multiple data using different strategies and methods that complement and strengthen each other, as also explained before. Qualitative methods play a crucial role in DBR. Qualitative methods generate rich data and can lead to deeper understanding of differing perspectives (Kuper et al. 2008). Qualitative data lend themselves to explore how and why complex phenomena occur and to study dynamic processes (Johnson & Onwuegbuzie 2004). But, whether qualitative data indeed contribute toward explaining complex phenomena is of course dependent on whether qualitative data are well analyzed and contribute toward explaining the findings and not just describing the themes that emerged from the data without explaining the relationships between the themes (Pope & Mays 2009).

The theory in practice

The context of the DBR studies

An example of DBR in which the authors of this article were involved dealt with the development of a teaching portfolio for stimulating teachers' professional development within the context of higher education (Tigelaar et al. 2006b, 2008). Modern theories on assessment, teachers' professional development, and teaching portfolios were used to develop a teaching portfolio prototype (Tigelaar et al. 2004b).

A teaching portfolio is an authentic assessment tool in which different measurement instruments are included, related to different teaching competencies, and in which feedback plays a crucial role. In line with constructivist views of learning and the corresponding need for alignment of learning and assessment, we wanted to develop a procedure for a teaching portfolio that was suitable for both formative and summative purposes in higher education. The portfolio should give teachers insight into their learning processes that occur as

Table 2. Overview of studies conducted to design, evaluation, and redesign a teaching portfolio.

Step 1: Analysis and redesign

1a: Development of a framework for teaching competence. Delphi study among educational experts. (Tigelaar et al. 2004a)

1b: Development of portfolio prototype. Interview study with portfolio experts. (Tigelaar et al. 2004b)

Step 2: Evaluation and redesign

2a: Exploration of teachers' reflections. Document analysis. (Tigelaar et al. 2006a)

2b: Teachers' interactions and their collaborative reflection processes during peer meetings. Video analysis. (Tigelaar et al. 2008)

2c: Participants' opinions about the usefulness of the portfolio. Interview study. (Tigelaar et al. 2006b)

Step 3: Analysis and redesign

3: Quality criteria and recommendations with regard to the portfolio assessment procedure. Literature study. (Tigelaar et al. 2005)

they are teaching and at the same time it should provide an appropriate basis for promotion decisions, i.e., decisions about the feasibility of pursuing a career in teaching on the basis of actual teaching performance. These combined objectives place high demands on the quality of the portfolio procedure and the judgment methods. Thus, the definition of the general problem that we addressed in this thesis was:

What form should a teaching portfolio take to be appropriate for both formative and summative purposes and how can the quality of such a portfolio be guaranteed?

The different steps involved in the design, evaluation and redesign of the portfolio are explained below. In addition, it is illustrated as to how the five characteristics of DBR above were applied in practice. It is also illustrated how theory testing and theory refinement and advancement of practice played a crucial role in this DBR study.

Continuous cycles of design, evaluation, and redesign of the portfolio

A cyclic process of defining theoretical underpinnings and design, evaluation activities, and redesign was the basis for the development of a portfolio procedure for both formative and summative teacher assessments. The research consisted of three main steps:

- (1) Analysis and redesign;
- (2) Evaluation and redesign; and
- (3) Analysis and redesign

In Table 2, we give an outline of the studies that were carried out.

The first analysis and redesign phase consisted of two substeps.

In step 1a, a framework for teaching competencies in higher education was developed and validated, which could be used as a starting point for appraisal of teachers' performance in higher education. This framework was developed by means of a Delphi study in which educational experts were asked to answer the question: "How important are the

following teaching competencies in each domain for an experienced teacher in higher education?"

In step 1b, a prototype for a teaching portfolio for teacher development and assessment was developed based on theoretical underpinnings. This prototype was tested against the opinions of portfolio experts after which the portfolio was redesigned.

The evaluation and redesign step consisted of steps 2a, 2b, and 2c. These studies investigated the effectiveness of the portfolio as an instrument for stimulating teachers' professional development, in particular for promoting teachers' reflections. In step 2a, we explored the formative function of the portfolio in a small-scale study in order to specify and revise the design. We analyzed portfolio assignments to identify illustrative examples of teachers' reflections on a variety of aspects of their functioning. In step 2b, we analyzed the quality of teachers' collaborative reflections during the process of working on the portfolio. Video recordings of peer meetings from the participating teachers were content analyzed in a fine-grained way. In step 2c, we examined enabling and disabling factors in relation to the stimulation of teachers' professional development by the portfolio. We also inquired whether teachers thought the portfolio was a good instrument for fostering professional development and for summative decisions about prospects for a career in education.

In step 3, we defined quality criteria or design guidelines for assessing teaching portfolios based on an analysis of a theoretical perspective on how to assess qualitative data. On the basis of the arguments presented in this study, design guidelines were developed for the organization of assessment procedures for teaching portfolios.

Finally, we returned to the general problem definition and aims and discussed what we learned from our design and development efforts as well as the implications with regard to the theoretical underpinnings and the practical implementation of teaching portfolios. Below, we explain how the characteristics of DBR as described above were applied in the different studies.

The research was carried out as a cyclic process of design, evaluation, and redesign. In step 1b, educationalists with ample experience with portfolio assessment were asked to evaluate the portfolio prototype. This resulted in several updates of the portfolio design guidelines. For example, The design guideline that the portfolio structure should be open and flexible and not be too prescriptive, in order to stimulate teachers' creativity and learning processes.

In step 2, a constant formative evaluation process was carried out as well. The portfolio design was updated based on an analysis of the content of teachers' reflections promoted by the different portfolio assignments (step 2a, Tigelaar et al. 2004b), a fine-grained analysis of the quality of teachers' collaborative reflections during the peer meetings (step 2b, Tigelaar et al. 2008), and teachers' opinions about the usefulness of the portfolio (step 2c, Tigelaar, et al. 2006b). This resulted, for example, in the guideline that differentiated portfolio assignments should be used in order to open the possibility to tailor a portfolio to a teacher's individual needs.

Step 3, which focused on specifying the guidelines for portfolio assessment, resulted in an update of the design

guidelines, too. On the basis of a literature study, we updated our theoretical underpinnings and design guidelines. We concluded that the most appropriate approach to portfolio assessment would be a hermeneutic, interpretative approach, because this would do justice to the qualitative nature of the information in portfolios. In addition, we concluded that criteria that are used in hermeneutic, interpretative research (Guba & Lincoln 1989) would be the most appropriate ones for building design guidelines for portfolio assessment processes. We redesigned our design guidelines in such a way that our formative and summative portfolio assessment procedure would be a process that is characterized by listening, dialog, and exchange of views with as little imbalance of power as possible. Some examples of design guidelines are: portfolio assessors should represent the different perspectives on teaching and have "prolonged engagement" with their candidates, i.e., be knowledgeable on the development process that candidates went through. Assessors should carefully consider all evidence in a portfolio, and challenge their interpretations with a peer assessor, searching for counterexamples in the portfolio evidence. Candidates should be given the opportunity to discuss about assessment criteria and interpretations about their portfolio and be given the chance to add information to their portfolio before final judgment is being made.

Authentic, real-life settings

The evaluation (steps 2a, 2b, and 2c) was carried out in an authentic, real-life setting. Five experienced teachers and their personal coaches were involved in the study, all medical doctors working at Maastricht Medical School. The teachers were selected for participation in the study on the basis of a proven interest in teaching, willingness to engage in professional development as a teacher, and the intention to pursue a teaching career. During the evaluation phase, the purpose of the teaching portfolio was primarily formative, although the option of future summative use was kept open. We asked the teachers to organize the portfolio according to the following roles of teachers: (Tigelaar et al. 2004a):

- the person as a teacher, expert on content knowledge, facilitator of learning processes, organizer, and scholar/lifelong learner.

In addition, teachers carried out six assignments:

- describe your teaching career history;
- describe critical incidents in your development as a teacher;
- design personal learning goals; select evidential materials;
- compose a profile of a good teacher; and
- compile a reflective portfolio.

After each assignment, the teachers alternately engaged in one of two types of social interactions: three meeting with their personal coach and three at a peer meeting. The social interactions were expected to reinforce the effects of the assignments through support and feedback, in-depth analysis, suggestions for improving practice, encouragement for teachers to reflect on their perceptions, and decisions in order to better understand their own teaching practice. The peer meetings were structured according to a systematic problem-solving

approach with minimal guidance by the facilitator. Global conversation protocols, i.e., guidelines for the dialogs between coaches and teachers were used to structure the meetings with personal coaches. The total time investment for participants was estimated to be 50 h.

Testing and refining theory and advancing practice

From the description above, it became clear that the studies revealed design guidelines that advanced practice, such as the portfolio structure should be open and flexible and not be too prescriptive, in order to stimulate teachers' creativity and learning processes. Below, we illustrate further how we refined our theoretical underpinnings (derived from theories about – teacher – learning) as a result of constantly testing those during the formative evaluations, illustrated by implications for design guidelines.

A first refinement of our theoretical underpinnings of the portfolio dealt with the purpose of the portfolio. We departed from the theoretical underpinning that in assessments, a high consistency between learning, instruction, and assessment is required. This is often referred to as the principle of "alignment," that is a high consistency between learning and assessment (Biggs 1996). Our viewpoint was that a portfolio should stimulate teachers' learning and professional development (formative function) and provide a basis for decisions about issues such as the possibility of an educational career (summative function). The dilemma is that this may cause undesirable interference of the two types of goals. The literature indicates that there may be negative backwash effects on teacher learning processes, for instance teachers might be only willing to present their strong points in relation to the criteria for summative assessment (Biggs 1996, 1999). This would seriously weaken the formative function, i.e.: teachers' learning processes and their professional development, because it would mean that areas that need improvement would remain underexposed to feedback and critical reflection. Since it is known that assessment drives learning (Longhurst & Norton 1997), assessment processes should be designed to provide meaningful learning experiences. Our studies demonstrated that the two aims can be combined in one portfolio procedure as long as the primary focus is on the learning of teachers and their professional development and the assessment process is performed in a careful and trustworthy manner. From our consultation of experts about the portfolio prototype (Tigelaar et al. 2004b), and on the basis of participants' opinions on the usefulness of the teaching portfolio (Tigelaar et al. 2006b), we refined our theoretical underpinning with respect to portfolio purpose. We argued that being able to work toward a summative decision from the start motivates teachers to work on their portfolio. Apparently, it is not realistic to expect that intrinsic motivation is a sufficient incentive for teachers to spend time on their portfolios and on reflecting on their teaching performance. Teachers also need to be convinced that the ultimate goals and profits make their efforts worthwhile. We therefore refined our design guidelines in such a way that a combined formative and summative portfolio procedure should be characterized by feedback and dialog about assessment criteria and an exchange of views

about good teaching in a climate with as little imbalance of power as possible.

Another refinement of our theoretical underpinnings dealt with the role of the coach. Should coaches only be support providers so as to maximize the safe climate of the coaching relationship or should coaches also have the summative function of assessor? Driessen et al. (2005) state that coaches should not be the formal assessors of the teachers they coach, because they are too closely involved with them. In our studies, we concluded that the two roles should not be strictly separated. According to the theoretical principle of prolonged engagement (Guba & Lincoln 1989), it is important that assessors should carefully consider all the longitudinal information that is available about candidates' teaching performance. This suggests that coaches, who have the most prolonged engagement with the candidates, should also be involved in the team of assessors who will discuss about what constitutes good teaching. In our research, we refined our theoretical principle about the coach role by combining the principle of prolonged engagement with constructivist views on learning and assessment, which stress learner participation and control (Birenbaum 2003; Segers 2004). From this viewpoint, our design guideline about the coach of the teacher who is being assessed was refined, arguing that coaches might provide additional context information about their candidates as a member of the team of assessors, which is otherwise comprised of other members, who are knowledgeable on teaching from different perspectives.

A third refinement of our theoretical underpinnings dealt with the difference between analytic and holistic assessments. Analytic assessment focuses on the application of precise criteria and standardized procedures, whereas holistic assessment is aimed at seeking to understand the whole in the light of all its constituents, constantly developing and testing hypotheses based on all assessment information. On the one hand, clearly defined analytic assessment criteria are needed to promote transparency and fairness, whereas on the other hand too much reliance on such criteria might turn portfolio assessment into a meaningless exercise of filling in blanks. The opposite holds for holistic portfolio assessment. From our studies, our theoretical underpinnings were refined in the sense that we came to the conclusion that meaningfulness and learning opportunities should take precedence over transparency and fairness. For our updates of the design guidelines, this means that criteria should be seen as guidelines that stimulate candidates to reflect on their functioning and encourage assessors to clarify their perspectives. Discussion and dialog about what constitutes good teaching will invite assessors to integrate different interpretations that emerge during the assessment of a portfolio into a meaningful holistic judgment. This means that assessment criteria should offer cues for consideration, without the need to pronounce judgment on precise criteria. The criteria, or steps for making interpretations, should be subject to regular revision.

Mixed-methods studies

We used a mixed-methods approach to gather our data, with an emphasis on qualitative methods. In step 1 of the research,

a combination of quantitative and qualitative methods was used. In order to develop a teaching competency framework, in step 1a, the Delphi-method was used; we developed a questionnaire which was filled out by educational experts, in different rounds. Subsequently, in step 1b, the development of a portfolio prototype based on theoretical principles, we carried out an interview study to determine to what extent the theoretical underpinnings were reflected in the prototype and to indicate any modifications they thought were required. Step 2, the evaluation phase, was carried out on a small scale. Therefore, quantitative methods were not very suitable. However, a mixed-methods approach was applied by combining different qualitative methods. This was done in order to gain insight into the learning processes and complex interactions that were stimulated by the portfolio. In step 2a, we carried out a document analysis, analyzing the contents of teachers' written reflections, based on their assignments. Subsequently, in step 2b, we carried out a fine-grained video analysis on the social interactions during the peer meetings and on the scope of teachers' collaborative reflection processes. In step 2c, we complemented these data with the opinions of teachers and their coaches about the usefulness of the portfolio structure and social interactions for stimulating their professional development. We carried out an interview study and analyzed findings using categories related to benefits, enabling and disabling factors in the portfolio. Finally, in step 3, in order to arrive at refined design guidelines for portfolio assessment, we carried out a literature study on methods for judging portfolios and quality criteria for guaranteeing the soundness of portfolio assessment processes.

Frequent interaction between designers, researchers, and practitioners

During the design process, frequent communication between designers, researchers, and practitioners was organized as part of the research process. The research and development process was carried out by a research team, consisting of four educational researchers, who also had expertise in educational development. The two authors of this article were members of that team. During the design process, we sought the opinions of both educational experts and practitioners from the field. For the Delphi study (step 1a), which was focused on developing a teaching competency framework, educational experts were selected who work in higher education. Subsequently, in step 1b, we developed and tested a prototype of a teaching portfolio for development and assessment purposes. The portfolio was submitted to the judgment of nine educationalists with ample experience with portfolio assessment. We investigated to what extent our theoretical choices were reflected in the prototype and which modifications were needed. In steps 2a, 2b, and 2c, practitioners were involved. Five experienced teachers and their personal coaches were involved in the study, all medical doctors. All these practitioners were involved in evaluating the usefulness of the portfolio, and their opinions were of great benefit for optimizing the portfolio design guidelines.

Pitfalls and challenges of DBR in general

Only of practical relevance for local situation

A disadvantage of conducting studies within an authentic real-life learning setting might be that they may result in interesting insights in a particular innovative learning environment within one local context, but these studies have limited added value for other contexts. In other words, the findings may not generalize to other contexts, because the study took place in a specifically designed context in which often a few students or teachers participate. On the other hand, several issues can be taken into account to stimulate a broader relevance than the local situation.

First of all, by grounding the design on theories and using theories to explain the findings.

Second, by including a detailed description of the context in the study, due to which readers can understand how or whether the findings can be applied to their own situation.

Third, by generating design principles.

Although the local context is very important in DBR, good DBR should yield guidelines that can be generalized across similar authentic innovative learning environments. Therefore, DBR requires testing a design across different settings (Design-Based Research Collective 2003). This testing across different settings does, however, not need to result in design guidelines that can be broadly applied to all contexts. At best, DBR studies should lead to defining design guidelines that make clear which characteristics are crucial for a particular intervention to be implemented in a specific context because of certain theoretical arguments, as argued by McKenney et al. (2006). So, it is not a matter of generating simple findings or design guidelines that can be generalized to all contexts that are broadly relevant for educational practice, as also argued by Regehr (2010).

Does not advance theories

Another disadvantage of conducting DBR might be that it is too much focused on designing and redesigning practice and too less focused on testing or refinement of theories. But, as stated before, good DBR is also aimed at testing or refining theories about the nature of complex learning environments (Design-Based Research Collective 2003). But, there are, of course, variations in DBR studies. Some DBR studies focus more on testing or proving theories (often called validation studies), whereas other studies focus more on using theories (often called development studies) (Nieveen et al. 2006). One major difference is that the former contribute toward the advancement of theories, whereas the latter contribute toward de-advancement of design guidelines. However, it is important to keep in mind that these latter studies often start with formulating design guidelines based on a review of the literature. In other words, these development studies also make use of theories. Furthermore, these development studies lead to the refinement of theories and testing whether these theories work in practice.

The involvement of the researcher might bias findings

In many DBR studies, the roles of researcher and designer are fulfilled by the same person. This might be difficult, because the researcher on the one hand must critically evaluate the design, but on the other hand is also the one who developed the design and must convince teachers or learners about the usefulness of the design. Of course, close interactions between designer, researcher, and practitioners in DBR are expected to lead to deep and meaningful insights, but it might also lead to a less critical evaluation of the design, e.g., because the researcher is convinced that the intervention does work or because the learners or teachers may find it difficult to criticize the design. In other words, the findings of the study might be more easily influenced by the researchers' biases. Triangulation of data sources and data methods is therefore of crucial importance (McKenney et al. 2006). The strengths of an additional method or data source can be used to overcome the weaknesses in another method and mixed methods can provide stronger evidence because of corroborating evidence (Johnson & Onwuegbuzie 2004). But, also, reflexivity is of crucial importance; i.e., the researchers should make the reader clear and give them information about their backgrounds and their lenses through which they conducted the study as well as what their pre-existing assumptions and expectations are, based on which readers can critically interpret the findings of the study (Bunniss & Kelly 2010).

DBR is risky, complicated, and time-consuming

Yes, indeed DBR is risky, complicated, and time-consuming, because of several reasons. One of the reasons is that the design must be based on theories and must be developed in close collaboration with researchers, designers, and teachers, and needs to be implemented in practice in close collaboration with teachers and learners. Thereafter, it can be evaluated and it should be redesigned. This cyclic process is risky, complicated, and time-consuming. It is risky because teachers or learners may withdraw and circumstances may change, from which the research design keeps changing. A research design that keeps changing is weak. Nevertheless, it is important that adjustments can be made to the design in close collaboration with the different stakeholders in order to design interventions that are useful in real practice (McKenney et al. 2006). Implementing innovations in practice often implies that changes must be made in order to meet the needs of all stakeholders. In other words, this is what happens in real practice. Another problem is that DBR is time-consuming because of its iterative process of design, evaluation, and redesign. Research programs nowadays often only consist of single projects that last for 3 or 4 years in which several studies need to be published and in which not much risks can be taken. Nevertheless, design studies should be conducted within long-lasting well-defined research lines or research programs in which different researchers work together and in which different studies are built upon each other. DBR researchers should work in teams in which they work together with various practitioners and various researchers with specific expertise on both quantitative and qualitative methodologies

on studies that are closely related to each other and can be built upon each other.

Dealing with pitfalls and challenges of DBR in practice

So far, we discussed the pitfalls and challenges of DBR in general. In the following section, it will be explained how we dealt with these pitfalls and challenges in our DBR studies that dealt with the development of a teaching portfolio as described above.

Practical relevance beyond the local situation

Although our portfolio was only tested in one setting, several measures were taken to optimize a broader relevance than the local environment in which the portfolio was tried out. First, by grounding the development process on theoretical rationales, and using these rationales to explain our findings, our studies were aimed to be of interest to a broader audience.

Second, by providing clear context information, and by being transparent in our analysis methods, we wanted to assist our readers in being able to judge applicability of our findings to their own contexts.

Third, by generating our design principles in an iterative process between theory and practice, and by showing the relationship between our design principles and our theoretical rationales, we aimed at inspiring the reader to carry out portfolio design processes in their own contexts, by building on our findings and other earlier studies.

A limitation to our study with respect to applicability of findings was the lack of detail in the design of the summative assessment procedure, the fact that the prototype was tested only twice, and the scope of the research, being particularly focused on higher education. Therefore, we recommended that a portfolio should first be tested across different settings to enable fine-tuning of all the elements in the portfolio process.

Advancing theories

The main focus of our studies was to contribute toward the advancement of design principles for portfolios by the use of theories (i.e., a development study) and not on testing or proving theories (i.e., validation studies) (Nieveen et al. 2006). However, we started with formulating design guidelines based on existing literature. Furthermore, we refined our theoretical rationales, based on our findings (for more on this issue, see section "testing and refining theory and advancing practice"). These theoretical rationales need to be tested in other settings.

Preventing bias of findings by involvement of the researcher

Similar to many other DBR studies, the roles of researcher and designer were fulfilled by the same persons. In order to prevent the findings of the study to be influenced by the researchers' biases, we used triangulation of data sources and data methods (McKenney et al. 2006). This was particularly done in the evaluation phase of the research, as explained before.

Furthermore, we applied principles of reflexivity by making them clear and providing information about the lenses through which we conducted our study as well as what our pre-existing assumptions and expectations were. This was done by clearly describing our theoretical rationales and by a transparent way of analyzing our results. This was also done by including measures for quality assurance of data analysis, such as having the authors challenge each others' interpretations after an independent analysis of the data, and by going back to the original data in case of disagreements about interpretations.

Reducing risks and complications: Preventing the research from being too time-consuming

The portfolio was tested in a real-life setting only once, and on a small scale, involving five teachers and their five coaches to prevent that the research would become too time-consuming. Although another cycle of testing in a real-life setting would have made the research stronger, we invested in using different data sources and methods in order to generate rich data on the impact of the portfolio on the teachers' professional development. Furthermore, while our research design and the different steps in the design process were clear from the beginning, it was possible to make adjustments to the design according to the results of the constant formative evaluations that characterized the research as a whole. This means that preliminary research questions were refined according to our results, and that the theoretical rationales to our research were fine-tuned along the process. In addition, the research and development team consisted of members with different kinds of expertise, so that different research perspectives could be combined. Finally, we used clear theoretical rationales which built on existing research lines in order to carry out our research.

Future applications and developments

DBR is, of course, only one approach among others that can be applied in medical education research. Which approach should be chosen is, of course, dependent on the specific research questions to be answered. DBR is an approach that fits well for educational research with a design focus (Reeves et al. 2005). It can be applied in many studies with a design focus, e.g., studies that focus on creating and testing innovative learning environments, introducing instructional technology in education, or designing assessment programs for innovative learning environments. DBR studies can be classified as clarification studies which seek to answer questions such as "How does it work?" and "Why does it work?" and studies that deepen and advance our understanding of medical education (Cook et al. 2008).

Below, two broad areas are described in which DBR could be applied, being the design of work-based learning environments and design of assessment programs.

Work-based learning

Research in the domain of work-based learning has revealed several insights about this learning environment; an environment that plays a dominant role in medical education. Work-based learning environments offer students many possibilities to interact with patients and medical experts or to participate in clinical practice (Doran et al. 2007; Teunissen et al. 2007). Many studies reported about the difficulties encountered within this environment, such as a lack of supervision of students and lack of feedback (Grant et al. 2003; Daelmans et al. 2004; Hofman & Donaldson 2004; Dolmans et al. 2008). This is a big problem, because supervision is a key success factor for effective student learning at the workplace (Wimmers et al. 2006; Kilminster et al. 2007). Studies within this field have, for example, led to the development and implementation of instruments that can be used to evaluate the quality of the clinical learning environment (Boor et al. 2007) or to evaluate the performance of clinical supervisors (Stalmeijer et al. 2008). Recently, it has led to the implementation of longitudinal attachments in undergraduate medical training programs to increase student continuity with patients and supervisors (Wamsley et al. 2009). It has also led to the development of a clinical teaching model, based on theories of cognitive apprenticeship, describing a three-step clinical teaching approach, that provides guidelines for teachers (Stalmeijer et al. 2010). So, various initiatives have been undertaken aimed at optimizing work-based learning. But, DBR studies could play a good role in further investigating and improving work-based learning. DBR studies could be conducted in which clinical staff, together with designers and researchers, redesign the workplace learning environment in such a way that clinical staff supervise students more efficiently. But, improving workplace learning requires not only the introduction of new tools or designs in educational practice, but also a cultural change, commitment, and involvement from all participants at the workplace. This can be enhanced by conducting DBR in which both clinical staff, students, designers, and researchers closely collaborate. DBR could contribute toward richer understanding of work-based learning environments, contribute toward apprenticeship learning theories, and also toward further improvement of work-based learning.

Assessment

Assessment is an area that has been investigated intensively during the past 20 years and has led to major changes in assessment practices. It has led to introducing objective structured clinical examination, portfolios and more recently all kinds of authentic work-based assessments tools (Norcini 2003). Still, little is known about how assessment can drive students toward deep learning or enhance learning. How can we insure that assessment and learning are not separate, but integrated? Shephard (2000) argues that we should create a learning culture where students and teachers would have a shared expectation that assessment should enhance learning. Teachers and students should look to assessment as a source of insight and help instead of a source for rewarding and punishing. Schuwirth and van der Vleuten (2004) also argue

that assessment and learning or teaching should be congruent with each other and that we should involve students in our assessments, e.g., by means of introducing peer and self-assessments. Another recent plea is to use a programmatic approach about assessment in education which consists of a combination of different assessment methods and data that drive student learning and can be aggregated for making decisions (Van der Vleuten & Schuwirth 2005; Van der Vleuten et al. submitted). This requires research in actual practice in which learning and assessment are intertwined. DBR might contribute toward a richer understanding of how assessment can enhance student learning and contribute toward theory building. In addition, it might move us toward further improvement of assessment that services learning, because researchers, teachers, and students within DBR could share ideas and redesign learning environments in which assessment and learning are intertwined.

Future developments

Medical education research should contribute toward testing and refinement of theories that explain why something works or does not work under particular circumstances. Theories give different “lenses” through which to look at complicated issues. Theories broaden our understanding of situations and can be applied in practice (Reeves et al. 2008). DBR can contribute to bridge building between theory and practice. A close collaboration and interaction between researchers, designers, and practitioners in DBR is crucial to obtain deeper insight in such complex settings as medical education. DBR uses mixed methods and triangulates different sources of information that explain why an intervention with a particular aim is successful under which circumstances and in which context.

Conclusions

Medical education research has grown enormously over the years, but does not sufficiently contribute toward advancing theories. It is argued in this Guide that DBR is a fruitful approach in design research that on the one hand contributes toward studies that are highly relevant for practice and on the other hand contribute toward theory testing and refinement. DBR is conducted in authentic real-life settings where learning normally takes place. Furthermore, in DBR, researchers, designers, and practitioners (teachers and students) closely interact with each other and share their ideas. Both characteristics contribute to the usefulness of this type of research for educational practice. Good DBR also leads to testing and refinement of theories and focuses on clarifying why a specific design with a specific aim does work in a specific context and as such contributes toward the advancement of theories and design principles.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

DIANA DOLMANS, PhD, Associate Professor, working at the Department of Educational Development and Research and the School of Health Professions Education (SHE), Maastricht University, the Netherlands.

DINEKE TIGELAAR, PhD, Assistant Professor at the Leiden University Graduate School of Teaching (ICLON), Leiden University, the Netherlands.

The DBR studies on the design of a teaching portfolio as described in this article were conducted by Dineke and resulted in her PhD thesis entitled “Design and Evaluation of a Teaching Portfolio” at the time Dineke was employed at Maastricht University. Diana was involved as her supervisor. They both wrote this guide together.

References

- Albert M, Hodges B, Regehr G. 2007. Research in medical education: Balancing service and science. *Adv Health Sci Educ Theory Pract* 12:103–115.
- Badley G. 2003. The crisis in educational research: A pragmatic approach. *Eur Educ Res J* 2(2):296–308.
- Barab S, Squire K. 2004. Design-based research: Putting a stake in the ground. *J Learn Sci* 13(1):1–14.
- Berliner DC. 2002. Educational research: The hardest science of all. *Educ Res* 31(8):18–20.
- Biggs J. 1996. Enhancing teaching through constructive alignment. *High Educ* 32:347–364.
- Biggs J. 1999. Teaching for quality learning at university. Buckingham: SRHE Open University Press.
- Birenbaum M. 2003. New insights into learning and teaching and their implications for assessment. In: Segers M, Dochy F, Cascallar E, editors. *Optimising new modes of assessment: In search of qualities and standards*. Boston, London: Dordrecht. pp 13–36.
- Boor K, Scheele F, Vleuten CPM, Scherpbier AJJA, Teunissen PW, Sijsma K. 2007. Psychometric properties of an instrument to measure the clinical learning environment. *Med Educ* 41:92–99.
- Brown A. 1992. Design experiments: Theoretical and methodological challenges in creating complex interventions. *J Learn Sci* 2:141–178.
- Bunniss S, Kelly DR. 2010. Research paradigms in medical education research. *Med Educ* 44:358–366.
- Collins A, Joseph D, Bielaczyc K. 2004. Design research: Theoretical and methodological issues. *J Learn Sci* 13(1):15–42.
- Cook DA, Bordage G, Schmidt HG. 2008. Description, justification and clarification: A framework for classifying the purposes of research in medical education. *Med Educ* 42:128–133.
- Daelmans HEM, Hoogenboom RJI, Donker AJM, Scherpbier AJJA, Stehouwer CDA, van der Vleuten C. 2004. Effectiveness of clinical rotations as a learning environment for achieving competence. *Med Teach* 26(4):305–312.
- Design-Based Research Collective 2003. Design-based research: An emerging paradigm for educational inquiry. *Educ Res* 32(1):5–8.
- Dolmans DHJM, Wolfhagen IHAP, Heineman E, Scherpbier AJJA. 2008. Factors adversely affecting student learning in the clinical learning environment: A student perspective. *Educ Health* 20(3):e1–e10.
- Dornan T, Boshuizen H, Kind N, Scherpbier A. 2007. Experience-based learning: A model linking the processes and outcomes of medical students' workplace learning. *Med Educ* 41:84–91.
- Driessen EW, Van der Vleuten CPM, Schuwirth LWT, Van Tartwijk J, Vermunt J. 2005. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Med Educ* 39(2):214–220.
- Eva K. 2009. Broadening the debate about quality in medical education research. *Med Educ* 43:294–296.
- Grant J, Kilminster S, Jolly B, Cottrell D. 2003. Clinical supervision of SpRs: Where does it happen, when does it happen and is it effective. *Med Educ* 37:140–148.
- Gruppen LD. 2008. Is medical education research ‘hard’ or ‘soft’ research? *Adv Health Sci Educ Theory Pract* 13(1):1–2.
- Guba EG, Lincoln YS. 1989. Fourth generation evaluation. London: Sage.

- Hofman KG, Donaldson JE. 2004. Contextual tensions of the clinical environment and their influence on teaching and learning. *Med Educ* 38:448–454.
- Johnson RB, Onwuegbuzie AJ. 2004. Mixed methods research: A research paradigm whose time has come. *Educ Res* 33(7):14–26.
- Kelly A. 2004. Design research in education: Yes, but is it methodological? *J Learn Sci* 13(1):115–128.
- Kember D. 2003. To control or not to control: The question of whether experimental designs are appropriate for evaluating teaching innovations in higher education. *Assess Eval High Educ* 28(1):89–101.
- Kilminster S, Cottrell D, Grant J, Jolly B. 2007. AMEE Guide No. 27: Effective educational and clinical supervision. *Med Teach* 29:2–19.
- Kuper A, Reeves S, Levinson W. 2008. Qualitative research. An introduction to reading and appraising qualitative research. *BMJ* 337:a288, doi: 10.1136/bmj.a288.
- Longhurst N, Norton LS. 1997. Self-assessment in coursework essays. *Stud Educ Eval* 23:319–330.
- McKenney S, Nieveen N, van den Akker J. 2006. Design research from a curriculum perspective. In: van den Akker J, Gravemeijer K, McKenney S, Nieveen N, editors. *Educational design research*. London: Routledge. pp 67–90.
- Monrouxe L, Rees CE. 2009. Picking up the gauntlet: Constructing medical education as a social science. *Med Educ* 43:196–198.
- Nieveen N, McKenney S, van den Akker J. 2006. Educational design research: The value of variety. In: van den Akker J, Gravemeijer K, McKenney S, Nieveen N, editors. *Educational design research*. London: Routledge. pp 151–159.
- Norcini JJ. 2003. Work based assessment. *BMJ* 326(7392):753–755.
- Norman G. 2007. Editorial – how bad is medical education research anyway? *Adv Health Sci Educ* 12:1–5.
- Pope C, Mays N. 2009. Critical reflections on the rise of qualitative research. *BMJ* 339:737–739.
- Reeves S, Albert M, Kuper A, Hodges BD. 2008. Qualitative research. Why use theories in qualitative research? *BMJ* 337:a949, doi: 10.1136/bmj.a949.
- Reeves TC, Herrington J, Oliver R. 2005. Design research: A socially responsible approach to instructional technology research in higher education. *J Comput High Educ* 16(2):96–115.
- Regehr G. 2010. It's not rocket science: Rethinking our metaphors for research in health professions education. *Med Educ* 44:31–39.
- Segers M. 2004. Assessment en leren als een twee-eenheid: Onderzoek naar de impact van assessment op leren. [The dyad of assessment and learning: A study of the impact of assessment on learning.] Inaugural address given at the acceptance of professorship in Pedagogics, Educational Sciences in particular, at Leiden University. Leiden: Leiden University.
- Schiffedercker K, Reed VA. 2009. Using mixed methods research in medical education: Basic guidelines for researchers. *Med Educ* 43:637–644.
- Schuwirth LW, van der Vleuten CP. 2004. Changing education, changing assessment, changing research? *Med Educ* 38(8):805–812.
- Shephard LA. 2000. The role of assessment in a learning culture. *Educ Res* 29(7):4–14.
- Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Muijtjens AMM, Scherpbier AJJA. 2008. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 30:e272–e277.
- Stalmeijer R, Dolmans D, Wolfhagen I, Muijtjens A, Scherpbier A. 2010. The Maastricht Clinical Teaching Questionnaire (MTCQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 85(11):1732–1738.
- Teunissen PW, Scheele F, Scherpbier AJJA, van der Vleuten CPM, Boor K, van Luijk SJ, van Diemen-Steenvoorde JAAM. 2007. How residents learn: Qualitative evidence for the pivotal role of clinical activities. *Med Educ* 41:763–770.
- Tigelaar DEH, Dolmans DHJM, Wolfhagen HAP, Van der Vleuten CPM. 2004a. The development and validation of a framework for teaching competencies in higher education. *High Educ* 48:253–268.
- Tigelaar DEH, Dolmans DHJM, Wolfhagen HAP, Van der Vleuten CPM. 2004b. Using a conceptual framework and the opinions of portfolio experts to develop a teaching portfolio prototype. *Stud Educ Eval* 30(3):305–321.
- Tigelaar DEH, Dolmans DHJM, Wolfhagen HAP, Van der Vleuten CPM. 2005. Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Stud High Educ* 30(5):595–610.
- Tigelaar DEH, Dolmans DHJM, De Grave WS, Wolfhagen HAP, Van der Vleuten CPM. 2006a. Portfolio as a tool to stimulate teachers' reflections. *Med Teach* 28(3):277–282.
- Tigelaar DEH, Dolmans DHJM, De Grave WS, Wolfhagen HAP, Van der Vleuten CPM. 2006b. Participants' opinions on the usefulness of a teaching portfolio. *Med Educ* 40(4):371–378.
- Tigelaar DEH, Dolmans DHJM, Meijer PC, de Grave WS, van der Vleuten CPM. 2008. Teachers' interactions and their collaborative reflection processes during peer meetings. *Adv Health Sci Educ Theory Pract* 13(3):289–308.
- Van den Akker J, Gravemeijer K, McKenney S, Nieveen N., editors. 2006. *Educational design research*. London: Routledge.
- Van der Vleuten CPM, Schuwirth L. 2005. Assessment of professional competence: From methods to programmes. *Adv Health Sci Educ* 1(1):41–67.
- Van der Vleuten CPM, Schuwirth LWT, Driessen E, Dijkstra J, Tigelaar DEH, Baartman LKJ, Van Tartwijk J. submitted. A model for programmatic assessment fit for purpose.
- Wamsley MA, Dubowitz N, Kohli P, Cooke M, O'Brien BC. 2009. Continuity in a longitudinal out-patient attachment for year 3 medical students. *Med Educ* 43:895–906.
- Wimmers PF, Schmidt HG, Splinter TAW. 2006. Influence of clerkship experiences on clinical competence. *Med Educ* 40:450–458.