Research design and the predictive power of measures of self-efficacy

Beverley Moriarty
Charles Sturt University

The purpose of this enquiry was to examine how research design impacts on the predictive power of measures of self-efficacy. Three cautions for designing research into self-efficacy drawn from the seminal work of Albert Bandura (1986) and a further caution proposed by the current author together form the analytical framework for this enquiry. For illustrative purposes, the analytical framework was applied to one stage of the author's Mathematics for Initial Teacher Education Students (MITES) project, whose goal was to determine the effects on self-efficacy of an intervention that aimed to increase levels of mathematics competence among initial teacher education students. The findings of the present enquiry have implications for designing, reporting, and determining the credibility and contribution of research into self-efficacy and for making meaningful comparisons between studies, including those whose results differ from each other or from what might be expected.

Introduction

Self-efficacy relates to how well a person judges that he or she will be able to perform particular tasks. Alberta Bandura’s (1977, 1986) seminal research into self-efficacy continues to be the most quoted in the field. A recent search of data bases indicates that self-efficacy theory is still applied to a range of disciplines including medicine, sociology, and education, the latter representing different levels of learning and a wide selection of subject areas, such as mathematics, literacy, social studies, and the arts. The research that is given particular focus in this paper is one stage in the Mathematics for Initial Teacher Education Students (MITES) project (Moriarty, 2008, 2011). The purpose of the MITES project is to investigate changes in initial teacher education students’ levels of self-efficacy in terms of their confidence to solve mathematical problems and their confidence to teach others to solve mathematical problems as they undertake interventions or programs that aim to increase their levels of achievement in mathematics. The present enquiry applies a four-point analytical framework to one stage of the MITES project. Its purpose is to examine how research design impacts on the predictive power of self-efficacy measures and subsequently how the results of research into self-efficacy that employs a quantitative strategy can contribute credibly to knowledge in the field.

Self-efficacy is important because it impacts on degrees of effort that individuals are prepared to invest in their learning, which then affects achievement. Self-efficacy is concerned with the confidence that individuals have in their ability to reach goals as a consequence of their actions (Hemmings & Kay, 2009). Initial teacher education students who will be required to teach mathematics to primary-level students during their pre-service teaching practice and later as graduates to students in their own classes need to be confident that they can solve mathematical problems and confident that they can teach others to solve mathematical problems. Regardless of how well initial teacher education
students achieved in mathematics at the high school level it is usually some years since these students solved mathematical problems at the primary level. Many initial teacher education students therefore simply need the opportunity to revisit primary level mathematics in order to re-familiarise themselves with the basic concepts and to improve their self-efficacy to solve problems and to teach others to solve the same problems.

The next part of this paper looks more closely at self-efficacy, how it is defined, the importance of accuracy in self-efficacy judgements and how self-efficacy is related to attributions of failure. The paper also examines the intersection between self-efficacy theory and research design and presents an analytical framework that is used to examine the design of the MITES project. Discussion then relates more broadly to how the analytical framework can be used to design and determine the value and credibility of studies into self-efficacy in relation to their capacity to contribute to theory and practice.

Self-efficacy theory

Self-efficacy theory is concerned with one’s level of confidence to perform tasks successfully and the factors that influence the development or loss of that confidence. There are compelling reasons why self-efficacy should be defined and measured in specific rather than global terms (Bandura, 1986). Bandura argued that specific measures of self-efficacy are better predictors of levels of subsequent performance than are global measures of self-efficacy, which can conceal underlying differences between separate dimensions of the concept.

Mathematics is one subject area where this point can be easily appreciated. For example, mathematics consists of a number of strands, across which people do not always demonstrate equal levels of confidence and achievement. For example, logic suggests that the level of confidence that one has in being able to solve problems involving fractions may be quite different from one’s level of confidence in solving mathematical problems relating to space. Thus some people may have similar levels of self-efficacy (and achievement) across these areas but others may differ markedly. A person who has high self-efficacy in one or more of these areas and low in others might find it particularly difficult to make global judgments relating to his or her confidence to solve mathematical problems. In such cases global measures of the concept can be elusive. It would also be difficult to determine whether global measures were more indicative of a perceived average level of confidence across the different areas or whether more weight was placed on what one perceived to be areas of strength or areas of difficulty. Alternatively, the most recent experiences with mathematical problem-solving could be the most influential. The unreliability of such global measures and the ensuing difficulty of making comparisons across cases on the basis of an independent variable illustrate the logic of Bandura’s (1986) determination regarding the relative value of specific versus global measures of self-efficacy.

Researchers often ask participants to undertake tasks parallel to the items on self-efficacy scales and these tasks need to be just as specific as the corresponding self-efficacy items.
If, for example, participants are asked how confident they would be in solving mathematical problems using decimals, then the associated tasks on an achievement test would also require them to solve problems relating to decimals. This is what Bandura (1986) referred to as the criterial task. When self-efficacy is measured more globally it is difficult to predict how well participants will perform on very specific tasks. When the self-efficacy questions relate to decimals there is more chance of predicting how well participants will perform in that same area.

Bandura (1986) also maintained that the predictive power of measures of self-efficacy would be higher when there was little delay between the measurement of self-efficacy and the performance and measurement of the related criterial tasks. This is also logical, even if the assessment of the psychometric properties of the self-efficacy scales indicates that the scales have high levels of internal reliability. With increasing passage of time between the two events, that is, the completion of the self-efficacy scales and the performance of the related tasks, there is more chance that other factors might intrude to affect performance on the tasks. For example, participants may find that they are distracted by intervening events or thoughts or that they experience environmental or physiological factors that have an impact on their performance after they have completed the self-efficacy scales.

Bandura (1986) proposed four main influences on self-efficacy. These influences are one’s prior experiences in the area of interest, observations of others completing the same or similar tasks, feedback, and encouragement received relating to the execution of a task and participants’ emotional states. Participants who know that they solved mathematical problems involving addition, subtraction, and multiplication with ease in the recent past would be likely to predict with confidence that they could do so again. Observing others execute tasks similar to what they will be asked to complete also leads people to draw on their confidence in their own past performance in the related area to judge how well they think that they could perform. If, for example, a colleague were observed solving multiplication problems slightly more difficult than what the observer had previously attempted, drawing on one’s prior experience and success would also help to predict the chances of being able to be as successful as their colleague in completing the more difficult examples. Receiving encouragement and feeling positive about the prospect of undertaking a task can also help one to be confident. In fact, Bandura (1994) found that self-efficacy improves when one is positive but is lowered when one is despondent. Prior performance and what one thinks about it, however, was found by Bandura (1986) to be the most influential source of self-efficacy. It will later be shown how this point needs to be taken into account when designing studies into self-efficacy and the impact that it can have on the predictive power of measures of self-efficacy.

Another important element of self-efficacy theory relates to the accuracy of self-efficacy judgements and what might happen if judgements are more optimistic or less optimistic than would be expected given past performance levels and experience. Bandura (1986) distinguished between judgements that were slightly over-optimistic and those that were gross over-estimations. Being a little over-optimistic was regarded as favourable because additional effort might be expended if it is perceived that such effort might increase the chances of success. It is more difficult, however, to achieve goals that are significantly
Research design and the predictive power of measures of self-efficacy

beyond reach and self-efficacy theory regards such attempts as likely to be counterproductive. When students perceive that there is little chance of achieving a goal, they are less likely to expend the effort that might otherwise have enabled them to achieve that goal. This can occur even when students have the ability to succeed.

Yet another element in Bandura’s (1986) theory is concerned with to what one attributes failure and how this attribution is reflected in levels of self-efficacy. When faced with failure, people with high self-efficacy are more likely to attribute their failure to sources outside of themselves, such as the difficulty of the test. When they cite a lack in knowledge or skills as the reason for their failure, people with high self-efficacy in a particular domain regard themselves as being capable of developing the knowledge and skills needed to succeed (Bandura, 1993). Conversely, people with low self-efficacy are more likely to blame their own lack of ability for their failure and dwell on what they believe that they cannot do rather than focus their concentration and efforts on achieving success.

As Brady and Bowd (2005) found, practicing teachers who have low levels of self-efficacy in relation to teaching mathematics often attribute these low levels to lack of content knowledge and to anxieties that they developed as a result of their experiences in learning mathematics when they were children. This need to address lack of content is complemented by the findings of Brown’s (2012) research into non-traditional initial teacher education students’ mathematics self-efficacy levels. Brown argued that identifying approaches to teaching mathematics methods courses that increased non-traditional initial teacher education students’ levels of self-efficacy could have benefits for these students as they progressed through their degrees and later when they became practicing teachers. It is therefore also important to address problems where they exist at the pre-service level before initial teacher education students perpetuate the cycle by passing on their anxieties to their own students when they enter the teaching profession. It is equally important that research that aims to address these problems can reliably contribute to growing knowledge in the area.

The relationship between attributions of failure, levels of self-efficacy, effort, and achievement, therefore, appear clear. In fact it has been shown over a sustained period of time that levels of self-efficacy impact directly on achievement (Swackhamer, Koellner, Basile, & Kimbrough, 2007). Even so, Williams and Williams (2010) questioned whether there was sufficient empirical evidence to support the idea that self-efficacy and performance in the same area of endeavour have a reciprocal relationship. They questioned Bandura’s central premise of reciprocal determinism, that higher self-efficacy is associated with and promotes higher levels of performance which, in turn, lead to higher self-efficacy.

The influences of self-efficacy theory on research and practice have been wide-spread and extensive. When new insights are discovered, just as when the results of any new studies are reported, it is important to examine not only how these new discoveries may impact on the theory and enrich understanding, but whether the studies that are reported follow good practice in their design and execution. It is this type of examination that may
contribute to resolving questions such as the one raised by Williams and Williams (2010) regarding reciprocal determinism.

The paper now turns to the intersection between self-efficacy theory and research design. It will then be shown how elements of Bandura’s theory on self-efficacy provide the basis of an analytical framework to design studies in the area and how readers of research studies can use the framework to evaluate the credibility of studies that claim to make a contribution to theory and practice.

**Mathematics for Initial Teacher Education Students project**

The Mathematics for Initial Teacher Education Students (MITES) project has had a number of stages, with its prime purpose being to contribute to identifying and investigating interventions to improve the self-efficacy and mathematical problem-solving abilities of initial teacher education students. The hypothesis of there being a high correlation between being able to solve mathematical problems and one’s self-efficacy to solve those problems as well as to teach others to do the same has been supported by the evidence available through this study (Moriarty, 2011).

As Usher and Pajares (2008) noted, research into self-efficacy has predominantly employed quantitative strategies. The MITES project to date has had the same approach. One stage of the MITES project (Moriarty, 2008) involved 81 first year initial teacher education students across four campuses of a regional Australian university studying the same mathematics curriculum subject in their early childhood and primary education degrees. The subject also contained a strand relating to students’ own competence in solving mathematical problems. To pass the subject, students needed to meet the additional requirement of demonstrating at least 80% competence in their own problem solving in mathematics, as indicated in a competence test at either the beginning or the end of the semester. This benchmark was a Faculty requirement of all students in the course. After completing a competence test at the start of the semester, students were provided with detailed results showing them how well they scored across the areas of concepts, number, measurement, fractions, space, and chance and data. The 81 students who did not achieve at least 80% were offered additional classes and other support throughout the semester to help them gain the skills needed to reach an overall minimum of 80% competence on a parallel test of achievement at the end of the semester. Immediately prior to undertaking their competence tests, these 81 students agreed to complete scales to measure their self-efficacy levels in relation to their confidence to solve mathematical problems in these areas and to teach others to solve problems in these same areas.

The self-efficacy scales that were used for this stage of the study were almost exactly the same as for other stages of the study. The only differences were that the researcher worked with the lecturers in the mathematics subject to make any changes to the scales as a result of changes in the curriculum. The items on the self-efficacy scales matched the content covered in the subject and the content in the competence tests. The scales used in
Research design and the predictive power of measures of self-efficacy

the later stages of the study, therefore, differed from those constructed in the earlier stages by reflecting minor intervening changes in the curriculum. As with other stages of the MITES project, the psychometric properties of the self-efficacy scales were found to be high, with reliabilities ranging between 0.93 and 0.98 for this stage of the project. The scales were all quite distinct from each other because they each represented a different part of the mathematics syllabus, that is, concepts, number, measurement, fractions, space, and chance and data. For this reason it was not necessary to conduct factor analyses to explore or confirm the different factors.

Figure 1 shows some of the items that were used to measure students’ self-efficacy in relation to their own problem-solving.

Please indicate below how confident you would be in solving problems involving the following maths concepts.

Key:
1 = Not very confident at all
2 = Only just confident
3 = Reasonably confident
4 = Very confident
5 = Extra confident
6 = Super confident

<table>
<thead>
<tr>
<th>Concept</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decimal fractions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hundredths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Money</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Example of a self-efficacy scale from the MITES project

As with the other scales, this scale also had a parallel scale to measure students’ self-efficacy in relation to teaching others to solve problems in these same mathematical areas.

The two sets of related self-efficacy scales, which measured confidence to solve problems and confidence to teach others to solve problems in the specific areas within mathematics, were highly positively correlated (between \( r = 0.79 \) for confidence to solve and to teach others to solve mathematical problems related to knowledge of concepts and \( r = 0.90 \), related to space). Causality cannot be claimed but the correlations indicate that the scores co-varied, indicating that as self-efficacy to solve problems in mathematics increased so did self-efficacy in relation to teaching others to solve those same problems.

The relatively small sample size for this study was accommodated with the use of a repeated measures design, this being one of the advantages of this type of design. This was particularly important given the total number of items across the self-efficacy scales relative to the sample size. The final administration of the self-efficacy scales also included an item at the end that asked students to indicate their level of attendance at the additional
classes on offer throughout the semester. Students indicated whether they attended these classes on a regular basis, quite a few times or just once or twice.

The results of Within Subjects Contrasts indicated that the significant differences found between each test of self-efficacy pre- and post-test on each scale occurred irrespective of how often students attended the competence classes. As an example, the results for confidence to solve mathematical problems involving knowledge of concepts yielded higher estimated marginal means (4.36) for the post-tests compared to 3.23 on the pre-tests. These results were highly significant for all four multivariate tests (Pillai’s, Wilks’, Hotelling’s and Roy’s). Similar results were obtained for all other tests.

The fact that students’ levels of self-efficacy were significantly higher on all self-efficacy tests post-intervention regardless of the number of classes attended has implications for practice. Students knew before the competence classes started what content would be covered in each class. This means that they could choose to attend classes that would help them with those areas on the pre-test of competence where they did not perform well. While this area requires further investigation it may indicate that it makes little sense to require students to attend all of the competence classes when they may have performed differentially across the different areas of competence on the pre-test. This approach also affords the students a level of autonomy in deciding what is best for them; they know that they need to reach the minimum level of competence on the post-test if they did not achieve this level on the pre-test and that, unless they reach this level, they will need to repeat the subject the following year.

**Analytical framework and its application**

The three points outlined in the earlier section on theory relating to the importance of self-efficacy being measured in specific rather than global terms, the alignment between measures of self-efficacy and measures of performance on the criterial task and the importance of avoiding delays between the administration of self-efficacy scales and related tests of performance together comprise the three cautions that Bandura (1986) insisted needed to be followed in the design of studies into self-efficacy. The fourth caution proposed in the present enquiry is that measures of self-efficacy are administered before tests on the criterial task.

Completing the self-efficacy scales before completing tests on the criterial task would avoid the possibility that levels of self-efficacy would be based on the perceived difficulty of the test just completed. While Bandura argued that past performance and what one thinks about it has the most influence on levels of self-efficacy, undertaking a test immediately before completing the self-efficacy scales compromises the ability of the researcher to discern the degree to which participants’ relied on their thoughts about their past performances in the area as opposed to their thoughts about the difficulty of the test when completing these self-efficacy scales. Together, these four cautions provide an important intersection between theory and research design and form an analytical framework to examine elements in the design of studies such as the MITES project. The
framework is also useful for designing studies into self-efficacy such that the results of the research are likely to be reliable and credible.

There are several notable examples of a closer examination of Bandura’s (1986) cautions in relation to studies in which the level of correspondence between self-efficacy and performance was not high. Reviews of such studies by Pajares and Miller (1994) and Pajares (1996) are important landmarks because they brought into question results that might otherwise have been regarded as making contributions to the field of knowledge beyond which they were entitled. These reviews found that when self-efficacy was defined in global rather than specific terms or when there was a time delay between the measuring of self-efficacy and the measuring of performance on the criterial task, the results showed an unexpectedly low correspondence between self-efficacy and performance on the task with which it was associated, such as achievement in a particular area.

As has been indicated in the discussion above and the example of some items on the self-efficacy scale, the MITES study followed Bandura’s (1986) advice about ensuring that self-efficacy was defined and measured in specific rather than global terms. For example, items from the concept strand of mathematics related to decimal fractions, percentages, hundredths, ratios, and money. These areas were covered on the self-efficacy scales as well as on the competence tests.

The analysis of the competence test results showed variability of performance across the different areas, such that students could see that, while they may not have performed particularly well in one area, they may have performed better in others. Just as it was important for students to have this detailed information in order to know where to place their efforts throughout the semester and before the parallel test of achievement was given, it was equally important that their levels of self-efficacy were tested using the same differentiation. Global measures of self-efficacy could not capture the finer details of the differences that could occur on more specific measures. At best they might give an average across the different areas but the finer detailed differences would be lost.

With regard to the two main areas of self-efficacy (confidence in solving mathematical problems and confidence in teaching others to solve the same problems), defining self-efficacy in specific rather than global terms and constructing the competence tests to match, therefore, it could be seen that not only was Bandura’s (1986) first caution heeded, but that the results supported the idea that such differentiation was needed. At the same time, Bandura’s second caution, that the performance test (in this case the test of achievement) should be closely aligned with the self-efficacy scales was also heeded, such that there was a direct parallel between the items on the self-efficacy scales and the items on the achievement tests. This means that testing for the degree of correlation between the results on the different tests and scales was appropriate and meaningful. Without such close correspondence between the specificity on the self-efficacy items and the items on the achievement test, it would be difficult to draw conclusions about the extent to which the scores co-varied.
The administration of the self-efficacy scales and the tests of achievement as both pre- and post-tests was completed in that order and without delay. Thus students completed both sets of tests in the one sitting on each occasion. Ensuring that the self-efficacy scales were completed before the achievement tests meant that participants’ responses on the self-efficacy scales were not influenced by what they perceived to be the difficulty of the achievement tests. Avoiding delays between the administration of the self-efficacy scales and the related achievement test ensured that the chances of other factors that could have impacted on achievement results did not occur in the space between these administrations. As the scales relating to students’ confidence to solve mathematical problems and the scales relating to students’ confidence to teach problem solving in the same areas were administered at the same time (in succession and without intervening delay), and given that the reliabilities for each scale were high, it was appropriate to be able to conclude from the results of the correlations that the scores on these measures also co-varied, such that as self-efficacy in relation to solving mathematical problems increased, so did students’ confidence to teach problem solving in the same mathematical areas. Bandura’s (1986) third caution relating to minimising the delay between the administration of measures of self-efficacy and achievement if also applied to the two areas within self-efficacy means that it is possible to test hypotheses and draw conclusions that would not be appropriate had there been a time delay.

**Discussion**

Research typically reports the psychometric properties of the scales that it develops and uses to measure the construct of self-efficacy. It is essential but not sufficient that such scales have high levels of reliability and that these levels are reported. Studies also need to be designed such that they maximise the predictive power of the self-efficacy scales that they employ. To achieve this goal they need to ensure that self-efficacy measures are specific rather than global, that there is a close correspondence between the items on the self-efficacy scales and the items on the corresponding criterial task, such as an achievement test, and that there is little delay between the administration of the self-efficacy scales and the related tasks.

Apart from satisfying these three cautions that Bandura (1986) identified it is also argued that self-efficacy scales need to be completed before the corresponding criterial tasks. If the corresponding task is in the form of an achievement test then completing self-efficacy scales after the achievement test means that participants could be basing their self-efficacy responses on what they perceive to be the difficulty of the test rather than on their confidence to perform such tasks in the future. It is also known from Bandura’s theory about attributions of failure that high- and low-efficacious students who perceive that they have not done well in the test are likely to react in different ways. Highly-efficacious students are likely to attribute their failure to outside factors such as the difficulty of the test while students with low levels of self-efficacy are more likely to blame their performance on factors within themselves, such as low ability. A similar problem can also occur among high- and low-efficacious students who perceive that they performed well on the test, with the latter students likely to think that the test may have been easy.
These examples mean that if tests of competence are completed before the self-efficacy scales, not only does this provide an opportunity for students to base their self-efficacy responses on what they perceive to be the difficulty of the test but that students whose self-efficacy levels are at the more extreme ends may react differently. Thus, requiring achievement tests to be completed before the related self-efficacy scales has the potential to establish a complex set of interactions that impact on the extent to which the results on the self-efficacy scales are truly reflective of students’ levels of self-efficacy. The extent to which these interactions impact on the results would be difficult to determine and brings into question the findings of such studies regardless of how well the self-efficacy scales and achievement tests were constructed and regardless of the reported levels of reliabilities of those scales and tests.

In landmark studies, Pajares and Miller (1994) and Pajares (1996) drew attention to the importance of research design through their critique of research that questioned previous findings and appeared to make new insights. It is equally important, however, to apply the same level of scrutiny to research that appears to be consistent with the accumulation of findings of previous research. Research into self-efficacy that has design deficiencies but does not attract attention to this point either because insufficient information about the design of the study is provided in the research report or because the findings are not surprising are just as problematic. They may give the impression of making a contribution to the field that is then drawn into the body of established knowledge that may then influence later research conducted by novice or experienced researchers. While it is not proposed or believed that the problems identified by Pajares and Miller are endemic or even common, well-designed and carefully-reported research impacts positively on the predictive powers of measures of self-efficacy and gives confidence and direction both to future research as well as practice.

**Conclusion**

It is a logical conclusion from this discussion that researchers need to think carefully about the designs of their research into self-efficacy, particularly in terms of Bandura’s (1986) cautions and the additional caution that together formed the analytical framework for the present enquiry. Researchers also need to provide enough information about their research designs in the reports of their studies to assist readers to be able to judge for themselves the credibility of the findings presented and the potential contribution that the findings make to theory and practice. These recommendations are important regardless of whether the findings are consistent with what might be expected from prior research in the area or whether the findings contest what is already known.

In studies where the results contest the findings of previous research, the first step is to determine whether those studies were well designed. In particular, if the four elements of the analytical framework applied to the present enquiry are all accounted for, then these new findings need to be taken seriously as they may signal new directions that were not previously known. If the findings appear credible then further studies should be undertaken to determine whether the findings are replicated. If, however, the examination
of the design of such studies reveals deficiencies relating to one or more of the four elements, then there needs to be some debate regarding the reliability of the findings.

Determining the credibility of the findings of research into self-efficacy and the extent to which their contributions to the field are meaningful is facilitated when the research report details how the design of the study took into account Bandura’s cautions and the additional caution proposed by the current author. Together these cautions formed the elements of the analytical framework that was used in the present enquiry to examine one stage of the design of the MITES project, whose purpose is to determine the effects on self-efficacy of interventions developed to increase the mathematical competence of initial teacher education students.

A final point to note is that the current enquiry related to research into self-efficacy that used the historically predominant quantitative strategy. Future enquiries of this nature could consider how the analytical framework used here might apply in parallel ways to qualitative research in the area, particularly given Usher and Pajares’ (2008) suggestion that the inclusion of qualitative approaches to the study of self-efficacy could lead to a richer understanding of the sources of self-efficacy beliefs. This would be a welcome addition to later stages of the MITES project as well.

References


Dr Beverley Moriarty is a senior lecturer in the School of Teacher Education at Charles Sturt University. Her most widely-cited research is in the area of self-efficacy and learning environments. She has also published extensively with the Australian Traveller Education Research Team and in lifelong learning.
Email: bmoriarty@csu.edu.au