Chapter 14
Using an Assessment *for* Learning Framework to Support Pre-service Teachers' Mathematical Modelling Activities

George Ekol

Abstract Assessment for learning is discussed within the broader framework of formative assessments. The aim of the study is to understand the contributions of assessment for learning in a pre-service secondary teacher mathematical modelling course at a university in South Africa. A matched pairs design was adopted to analyse assessment data collected during the course, and data collected at the end of the course. Descriptive and inferential data analysis detected no statistically significant increase in the mean score from the formative phase ($M = 78.33$, $SD = 8.86$) and the scores obtained from the final assessment at the end of the course ($M = 81.52$, $SD = 10.97$), $t(62) = 1.728$, $p = .089$, $\eta^2 = 0.218$. The study contributes to research on various assessment approaches in pre-service mathematics education courses that include mathematical modelling and understanding their practical contributions to the learning gains at the end of the courses.

14.1 Introduction

14.1.1 *Mathematical Modelling*

Many countries worldwide have introduced mathematical modelling in their school curricula, in part due to the awareness that has been created over the last forty years through various platforms including the *International Conference for Mathematical Education* (ICME) and the *International Community of Teachers of Mathematical Modelling and Applications* (ICTMA) conferences and publications. Modelling is understood differently among various communities. In this study, I adopt Niss et al. and's (2007, p. 4) representation of a mathematical model, consisting of a domain

G. Ekol (✉)
Wits School of Education Parktown, University of Witwatersrand, Johannesburg, South Africa
e-mail: george.ekol@wits.ac.za

D of interest outside mathematics (extra-mathematical), a mathematical domain *M*, and a link from outside mathematics to the mathematical domain. Questions from outside mathematics that need to be understood are identified and linked to the mathematical domain, where elaborate mathematical treatment and inferences are made, the outcomes of which are then translated back to *D*. In *D*, interpretations and validations are made in response to the original question. The back-and-forth movement (modelling cycle) between *D* and *M* can be done according to the need, and as many times as possible until a satisfactory conclusion concerning the original question from *D* is reached. The whole process comprising of structuring *D*, to deciding upon a suitable mathematical domain *M* and a suitable mapping from *D* to *M*, to working mathematically within *M*, to interpreting and evaluating the conclusions with regard to *D* is the modelling process (Blum, 2015; Greefrath & Vorholter, 2016; Niss et al., 2007; see also detailed explanation in Wess et al., 2021, p. 6).

Figure 14.1 (Blum & Leiß, 2007) represents complex cognitive processes and the associated affective processes that students undergo when they engage in modelling tasks (see later an elaboration of modelling tasks). Students should be able to translate between reality-based problems into mathematical models and to work within the mathematical model to gain understanding of the problem. The ability of students to perform such modelling tasks is an indication of their modelling competence (Kaiser, 2007; Geiger et al., 2022). Promoting the students' ability to process real-world problems with mathematical tools is an important goal of modelling in school mathematics.

The seven sub-competencies in Fig. 14.1 have been elaborated in several research publications (e.g., Greefrath et al., 2013; Greefrath & Vorhölter, 2016; Wess et al., 2021). Descriptions of the modelling sub-competencies are briefly presented in Table 14.1.

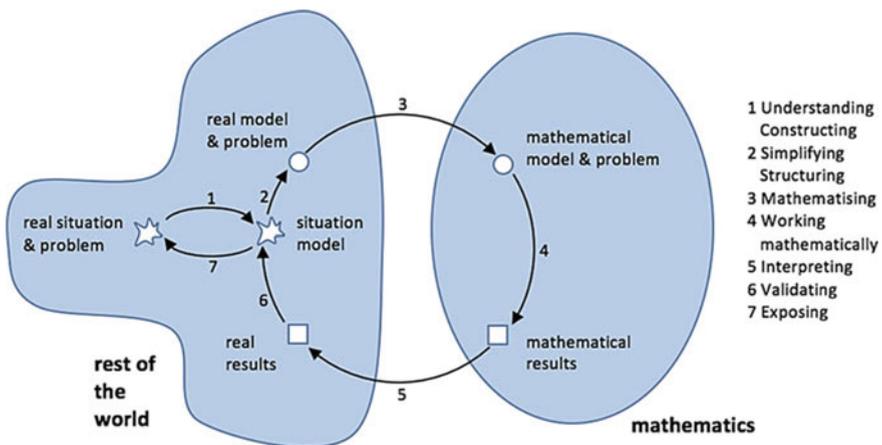According to Niss et al. (2007), mathematical modelling competence implies:

Fig. 14.1 Modelling cycle (Blum & Leiß, 2007, p. 221)

Table 14.1 Sub-competencies involved in modelling

Sub-competency	Description
Constructing	Students construct their own mental model from a given problem and thus formulate an understanding of the problem
Simplifying	Students identify relevant and irrelevant information from a real problem
Mathematizing	Students translate specific, simplified real situations into mathematical models (e.g., terms, equations, figures, diagrams, functions)
Working mathematically	Students work with mathematical methods in the mathematical model and get mathematical solutions
Interpreting	Students relate results obtained from manipulation within the model to the real situation and thus obtain real results
Validating	Students judge the real results obtained in terms of plausibility
Exposing	Students relate the results obtained in the situational model to the real situation, and thus obtain an answer to the problem

the ability to identify relevant questions, variables, relations or assumptions in a given real world situation, to translate these into mathematics and to interpret and validate the solution of the resulting mathematical problem in relation to the given situation, as well as the ability to analyse or compare given models by investigating the assumptions being made, checking properties and scope of a given model etc. (Niss et al., 2007, p. 12)

Supporting students to solve real-world problems using mathematical tools available to them is therefore a central goal of modelling in school curricula, for example in South Africa (DBE, 2011, p. 8). Blum (2015) conceptualized the modelling competence as being able to construct or adapt mathematical models by conducting process steps adequately and being able to compare and analyse different models. However, the challenge has remained in assessing students' sub-competencies in classroom environments when they solve different modelling tasks.

14.1.2   *Mathematical Modelling Assessment*

The debate about performance assessment and how it can be used in the classroom is ongoing among educational researchers, schoolteachers, and the mathematical modelling community globally. Many mathematical modelling related assessment frameworks have been developed over the years (e.g., Alagoz & Ekici, 2020; Besser et al., 2013; Rakoczy et al., 2017). Besser et al. (2013) investigated how assessing and reporting students' performances in mathematics can be arranged in everyday teaching in such a way that teachers are able to analyse students' outcomes appropriately. Their mathematical tasks focused on technical and modelling competencies of students. Besser et al. premised the study on the assumption that assessing, and reporting students' outcomes regularly would foster learning processes and

improve performance for the experimental group. However, they found no significant differences between a control group and an experimental group in a posttest. In another project, 'Conditions and Consequences of Classroom Assessment' consisting of four studies, Rakoczy et al. (2017) successively investigated the impact of formative assessment in mathematics instruction. The project comprised a survey study, an experimental study, an intervention study, and a transfer study to make the results applicable in educational practice through teacher training in formative assessment. Concerning the impact of teacher training on pedagogical content knowledge, Rakoczy et al. (2017) concluded that knowledge about formative assessment in competence-oriented mathematics instruction with a focus on mathematical modelling was significantly higher when teachers participated in training on formative assessment, compared to teachers who trained in general aspects of competence-oriented mathematics instruction and problem solving. A study by Alagoz and Ekici (2020) involved a mathematical modelling assessment approach designed to provide feedback regarding the performance of each learner and on the task itself. One benefit of Alagoz and Ekici's (2020) study was that it enabled the authors to identify professional development needs for teaching mathematical modelling and applications. For example, their data analysis indicated that teachers had difficulty in connecting between different concepts, with fewer teachers demonstrating mastery than other attributes, whereas problem solving was where they performed best. As making connections was the weakest aspect of mathematical modelling performance, Alagoz and Ekici (2020) proposed that more training and support were needed for teachers. To support interdisciplinary connections, the researchers recommended interdisciplinary professional development programmes where mathematics and science teachers can support each other to develop richer and more meaningful connections and interpretations with modelling and applications. Communications and representations as performance attributes were other areas where the teachers showed need for improvement as well. As these three studies reviewed show, assessment findings vary according to design objectives and context within which a study is situated. Nevertheless, lessons can be selected from one assessment setting and tried in another with modifications.

Since the 1980s, successive assessment criteria have been developed that incorporate the seven modelling sub-competencies in Table 14.1 (e.g., Berry & Masurier, 1984; Haines & Dunthrone, 1996; Hall., 1984; Hankeln et al.,. 2019; Hidayat et al., 2022; Houston, 2007; Izard et al., 2003; Leong, 2012; Penrose, 1978). The assessment criteria developed have, in general, favoured holistic assessment—assessing modelling. Micro-assessment of individual sub-competencies have been reported by Hankeln et al. (2019). Using psychometric models, Hankeln et al. (2019) showed that the sub-competencies of modelling, simplifying, mathematizing, interpreting, and validating, can be treated as separate dimensions, rather than being subsumed into a two-dimensional model, in which *simplifying* and *mathematizing,* as well as *interpreting* and *validating,* have been combined. Although much progress has been made, assessing modelling activities in pre-service courses and at the school level is still a big challenge.

The recent review of modelling research worldwide by Hidayat et al. (2022) revealed that the three dominant approaches used in assessing modelling competency are projects (50%); written tests (28%); and questionnaire (22%). Almost one-third of the published papers employed the qualitative approach as a method of data collection with the highest percentage of participants being pre-service teachers. Assessment involving project work seemed a preferred approach because modelling is commonly thought of as a collaborative process (Houston, 2007), so a comprehensive approach was seen as the best method to assess students' modelling competency. It is not surprising therefore that pre-service teachers dominated the papers because at the undergraduate level, some flexibility is assumed for them to complete project work on their own. However, flexibility at undergraduate or even graduate level, cannot be assumed if institutional cultural variations are considered up to the microlevel of timetabling or sharing the available teaching resources. In cases where timetables are fully booked for teaching other subjects, project work will be difficult to find time for. Hence, the written assessments have remained the more preferred approach. Yet, there are indeed other assessment approaches that can be incorporated without changing the existing structures of timetables. The question is, "When mathematical modelling is introduced into traditional courses at school or university, how should existing assessment procedures be adapted?" (Blum & Leiß, 2007, p. 23). One possibility is assessment *for* learning.

14.1.3   Assessment for Learning

Summative and formative assessments are two frequently used assessments in schools worldwide. Summative assessments (also sometimes referred to as assessment *of* learning) are types of assessments that are used to measure what students have learnt at the end of a unit for purposes of promoting a student to the next grade, or for certification after completing school. Assessment *for* learning (also often called formative assessment) is assessment that puts emphasis on the processes of teaching and learning and aims at actively involving students in those processes. Assessment for learning (AfL) also aims to build students' skills for self-assessment and helping them to understand their own learning, and to develop appropriate strategies for lifelong learning (OECD, 2008).

AfL also prioritizes the regulation of learning processes. The assumption is that with the regulation of processes, classroom assessment can be used to improve learning. Regulation involves four main processes: goal setting; monitoring of progress towards the goal; interpreting feedback derived from monitoring; and adjusting goal-directed action where adjustment is needed at the time (Allal, 2010). It is this orientation that is most often referred to when speaking of "formative assessment," but I use AfL with an emphasis on the process of learning than on the product of learning although both are important. Since AfL is also the orientation that forms the foundation of this study, it will be discussed in more detail.

Research and practice in classroom assessment emphasize similar regulatory goals and processes (e.g., Andrade & Heritage, 2018; McMillan, 2013; OECD, 2008). Defined as a process of collecting, evaluating, and using evidence of student learning to monitor and improve learning (McMillan, 2013), effective classroom assessment makes clear the learning targets, provides feedback to teachers and to students about where they are in relation to those targets, and prompts adjustments to instruction by the teachers to meet students' learning needs (Andrade & Heritage, 2018). Hattie and Timperley (2007) summarize this regulatory process in terms of three questions to be asked by students: (i) Where am I going? (ii) How am I going? and (iii) Where to next? The three questions are also asked by the teachers in reference to their students' learning. Starting with clear learning goals and task criteria, collecting and interpreting evidence of progress towards those goals and criteria, and finally acting by adjusting instruction or learning processes, the regulatory processes of AfL are implemented. Particular attention is placed on the third stage, the where next? The stage involves taking action to move students towards the learning goals (Andrade & Heritage, 2018). It involves drawing on feedback from the students to revise the learning activities (Wiliam, 2010).

Assessment for learning incorporates lifelong learning (OECD, 2008). Teachers using formative assessment approaches guide students towards developing their own "learning to learn" (p. 2) skills—being flexible and inquisitive about learning current ideas and methods of solving a problem—that are increasingly necessary as knowledge quickly becomes out of date in today's volatile information environment. Six key elements of AfL that emerged from studies conducted by OECD (2008) are: (i) A classroom culture that encourages interaction and the use of assessment tools; (ii) setting up of learning goals, and tracking of individual student progress towards those goals; (iii) use of varied instruction methods to meet diverse student needs; (iv) use of varied approaches to assessing student understanding; (v) feedback on student performance and adaptation of instruction to meet identified needs; and (vi) active involvement of students in the learning process. OECD highlights the tension between formative assessment and summative assessment. Summative tests—that is, large-scale national or regional assessments of student performance hold schools accountable for meeting the set standards. The consequences of such highly summative tests take up much of the resources that would otherwise be directed to supporting the assessing for learning.

AfL is also regarded as an avenue for improving student learning and enhancing their course achievements (Gan et al., 2019). The AfL movement has historical links with the Assessment Reform Group (Black & Wiliam, 1998) which proposed a distinction between *assessment of learning* for the purposes of grading and reporting, and *assessment for learning* which promotes providing information for both the student and the teacher to improve learning and to adjust teaching. AfL has been defined as: "part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning" (Klenowski, 2009, p. 264).

Building on socio-constructivist theories of learning, AfL puts the focus on what is being learned and on the quality of classroom interactions and relationships (Stobart,

2008), starting from the learner's existing knowledge, and emphasizes the need for active and responsible involvement of the learner and the value of developing metacognition (Black, 2015). AfL is also characterized by a process of continual interaction between teachers and individual learners, in which feedback provision and its acceptance and utilization are key elements (Black & Wiliam, 1998). By feedback 'acceptance and utilization', Black and Wiliam suggest that the student must act upon the feedback he or she has received from the teacher for the required change to materialize (Wiliam, 2011). The teacher-student interaction during the course unit is iterative in that a student's response provides additional information for the teacher to act upon and adjust teaching (Kennedy et al., 2008).

14.1.4 Modelling Activities

Modelling activities refer to the modelling tasks that the pre-service teachers are engaged with during the course. The activities require more preparation by the teacher than in traditional teaching approaches (e.g., Antonius et al., 2007). In modelling activities, teachers plan for, and students spend more time on substantial tasks. Depending on the teaching arrangement, the modelling activities by students include discussing mathematics with each other; exploring alternative solutions to a given task; choosing appropriate mathematical artefacts (e.g., sketches, graphs, formula) to use in solving a task; reasoning about the solution of a task; and checking strategies to ensure that the solution is valid (Antonius et al., 2007, p. 296). Overall, students take more responsibilities, and the teacher's role is to monitor the progress and intervene where such intervention would move the learning forward. The modelling activities framework proposed by Antonius et al. (2007) completely agrees with the frameworks of the assessment for learning discussed in this study.

The modelling tasks in this course are familiar curricula tasks of various lengths, based on mathematics and applications. The broad curriculum coverage of the course includes modelling with linear functions, modelling with polynomials, and modelling with exponential and power functions. The course is aimed at preparing secondary school teachers. Two examples of short modelling tasks with linear and quadratic functions follow:

Example 1. A property owner wants to fence a rectangular garden plot adjacent to a main road. The fencing next to the road must be strong and costs $5 per metre, but the fencing for the rest of the field costs just $3 per metre. The garden has an area of 1200 square metres. Find the garden dimensions that minimize the cost of fencing. If the owner has a budget of $600 to spend on fencing, find the range of lengths that she can fence along the road.

Example 2. A national soccer team plays in a stadium with maximum capacity of 60,000 fans. With a ticket priced at $10, the average attendance at the recent games has been 30,000 fans. A survey conducted to gain an understanding of ticket pricing and its links to games attendance revealed that for every dollar that

the ticket price was lowered, the attendance would increase by 4000 fans. What ticket pricing would maximise the revenue collection? (Adapted from Stewart et al., 2015).

These questions allow students to construct mental models of real situations. Also very important are the assumptions that students need to make in each case to simplify the problem. Experience shows that students tend to approach the above questions as purely mathematical problems, not modelling problems. For instance, many students ignore the role of assumptions in simplifying complex problems. Although a sketch would help a student to develop a mathematical model, some students are not used to drawing sketches. Solving tasks such as these relatively easy problems evoke nearly all the modelling sub-competencies in Table 14.1, with the obvious ones being, constructing, simplifying, mathematizing, working mathematically, interpreting, and validating.

I have used the terms "activities", "tasks", or "problem" in a broader sense to mean learning activities of varying difficulties that are assigned to the students during modelling. Such tasks require several or all steps of the modelling cycle to solve (Durandt et al., 2022). The use of "problem" in a more strict interpretation in problem solving (e.g., Schoenfeld, 2013; Lester, 2013, p. 248), has not been applied in this case.

14.2   The Study

Fourth year pre-service secondary mathematics teachers ($N = 63$) at a university in South Africa participated in the study. The selection of participants was purposive in that it targeted this group of students taking mathematical modelling in their mathematics content course. The students had already covered other mathematics content courses such as, algebra, functions, geometry, financial mathematics, probability and statistics, linear algebra, and calculus during the four years of their B.Ed. programme. Modelling is the last course that the students take to complete their mathematics content courses. Assessment *for* learning has been reported in numerous studies to offer opportunities for "high-performance, high equity [in] student outcomes, and for providing students with knowledge and skills for lifelong learning" (OECD, 2008, p. 5). Assessment scores obtained by students during coursework and one final assessment administered at the end of the course, were analysed using matched pairs *t*-procedures. For this study, five course assessments from the course for pre-service secondary teachers incorporating a variety of modelling tasks were assigned and graded throughout the course. The mean mark in the course assessments for each student constituted one set of measurement data for the assessment *for* learning. The second measurement data were obtained from the final written assessment at the end of the course. From those two data sets, the matched pairs *t*-procedures were applied to check if the difference between the two means was statistically significant, and if so, to what effect? Hence, the measurement of the effect size was also considered.

The research questions for the study were:

1. Is there a significant change in the pre-service teachers' mathematical modelling scores at the end of a teaching plan that applies an assessment for learning (AfL) framework?
2. What is the impact of such a change, if any?

14.2.1 Research Design

14.2.1.1 Matched Pairs Design

Matched pairs design compares two treatments. Pairs of participants that are as closely matched as possible are chosen and matched. Chance is used to decide which participant in a pair is allocated to the first treatment and who to the second (Fig. 14.2). A paired-sample *t*-test is used to measure whether the difference in the mean scores after two different treatments at various times on a pair is statistically significant. The basic assumption is that the difference between the two scores obtained for each subject is normally distributed. With a sample size of more than 30 cases, violation of this assumption if any, is considered not to be severe (Pallant, 2020).

Another situation calling for matched pairs is the so-called *before* and *after* observations (Moore et al., 2013) but on the same participant. That means each participant is his or her own pair. An individual is assessed several times during the course and the mean score is recorded. The same individual is also assessed at the end of the course. To compare the responses to the two 'treatments' before and after, the difference between the responses within each 'pair' is obtained. A response to treatment refers to the mean score that a student obtains during the *assessment for learning* phase as well as the score from writing the final assessment at the end of the course. The one-sample *t* procedures are then applied to the differences between the scores (Moore et al., 2013).
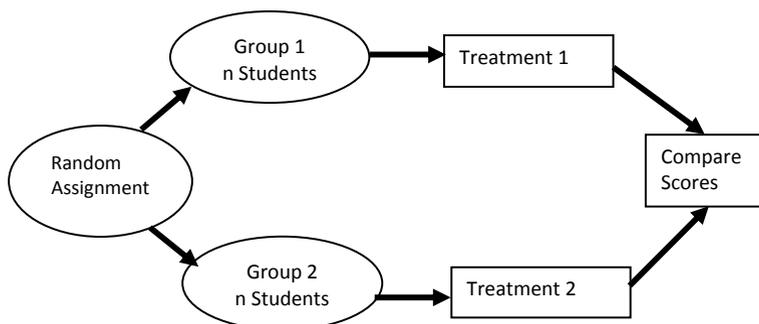

Fig. 14.2 Matched pairs design (Adapted from Moore et al., 2013, p. 236)

Taking μ to be the mean difference in scores (in the population of pre-service teachers) during the course and the score obtained in final assessment, the null hypothesis (H_0) tested was that the assessment *for* learning (AfL) has no effect on students' final grade. In other words, the difference between the scores obtained from the two assessments is zero. The alternative hypothesis (H_a) was that AfL makes a statistically significant contribution to the students' final score. Mathematically, the hypotheses are: $H_0 : \mu = 0$ and $H_a : \mu > 0$.

The parameter μ in matched pairs *t*-test procedure is the mean difference in the responses to the two treatments within matched pairs in the population. In this study, I have adopted the *before* and *after* observations and collected data in two phases. Phase one was the coursework duration where assessment *for* learning principles discussed earlier, were implemented. Phase two was a 3-h written assessment (examination) taken at the end of the course. The observed mean obtained from the assessment *of* learning and that from the written examination are compared using descriptive and inferential statistics in Sect. 3.

The matched pairs one-sample design was adopted mainly because the procedure fitted the one semester period that was available for the course. The design, for instance, did not require using a random procedure to split the class into two equal groups and teaching them separately. Adopting the procedure not only enabled uniformity in the content delivered to the students, but also it complied with the assessment guidelines provided by the institution such as having the end of course materials internally and externally moderated before assessing students on the materials. Finally, another important design principle that was implemented in the study was varying the assessment content given to the students during the course and at the end of the course.

14.2.1.2 Sample

Seventy-five (Male $= 56$, Female $= 19$) final year pre-service mathematics teachers enrolled in the eight-week long modelling course in 2021. Non-probability sampling was adopted where all final year pre-service secondary mathematics teachers registered in the modelling course automatically qualified to participate in the study. However, 12 students did not have complete data, so they were excluded from the analysis leaving 63 (Male $= 44$, Female $= 9$) cases.

14.2.2 Data Gathering and Analysis

The data consisted of two sets: one set was collected based on assessment *for* learning principles (as discussed in Sect. 1.3). Students' active involvement in the learning activities through interactions (OECD, 2008) was prioritized. Interactions included students sharing their solutions with peers and with the teacher on different platforms; teacher follow up with individual students; providing feedback to the group while

also attending to specific individual needs; varying the assessment methods such as asking students to present their solutions in words, in graphical format, and to present their solutions to peers in class. Finally, the course facilitators ensured that each student responded to the feedback given to them at different stages during the course. A total of five AfL assessments were completed and graded during the course, and one final assessment written at the very end of the course was also graded.

The mean scores obtained from the five assessments for learning (AfL) constituted the first measurement data set (T1) for each student. The scores obtained from the end of course assessment constituted the second measurement data set (T2) for each student. For the AfL framework to have contributed significantly to the pre-service teachers' learning gains, the following four assumptions (A1–A4) were tested using the quantitative data.

(i)   **A1**: The mean score obtained from the AfL phase is higher than the middle score of 50%. The 50% was arbitrarily chosen as a reference mark, but also used in the study as pass mark.
(ii) **A2**: The mean score in the final assessment is higher than the mean score obtained from the AfL sessions.
(iii) **A3**: The difference between the mean scores in (ii) and (i) for each student, is normally distributed.
(iv) **A4**: There is no statistically significant difference between the mean scores obtained in the AfL phase and in the final assessment. Alternatively, there is a statistically significant difference in the mean scores obtained in the AfL phase, and the final assessment. If the latter is true, then we conclude that the AfL contributed significantly to the pre-service teachers' learning gains at the end of the modelling course.

14.3   Results

To answer the research questions, quantitative data were analysed using IBM SPSS 28 software to check the four assumptions. The findings related to the four assumptions are now presented.

A1: The mean scores from AfL assessments are above pass mark: The mean score from AfL scores was found to be 78% (SD = 8.86, N = 63). The distribution of scores was reasonably normal (Fig. 14.3) for the mean to be used as a unit of measurement. Moreover, with a sample size of 63 (N > 30) cases, the normality requirement was not considered as a serious threat to the mean being used as a unit of measurement (Pallant, 2020).

A2: The mean score in the final assessment is higher than the mean from the AfL: The mean score obtained by the pre-service teachers in the final assessment was 81.52% (SD = 10.97, N = 63) showing a higher mean than that obtained during the course. Hence, assumption (ii) is also satisfied. As in the formative assessments, the final scores are also reasonably normally distributed (Fig. 14.4).
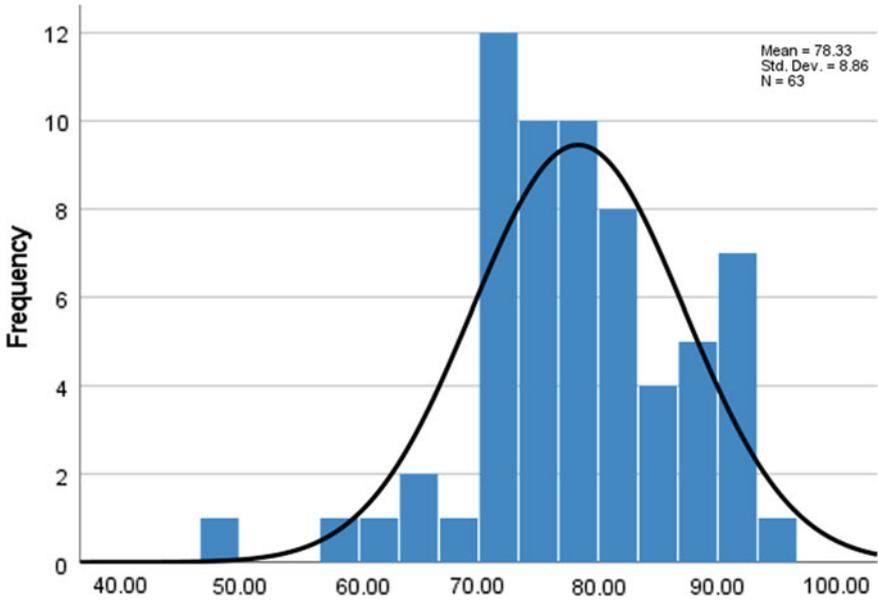
Mean = 78.33
Std. Dev. = 8.86
N = 63

Fig. 14.3 The distribution of the mean of assessment scores during the AfL phase

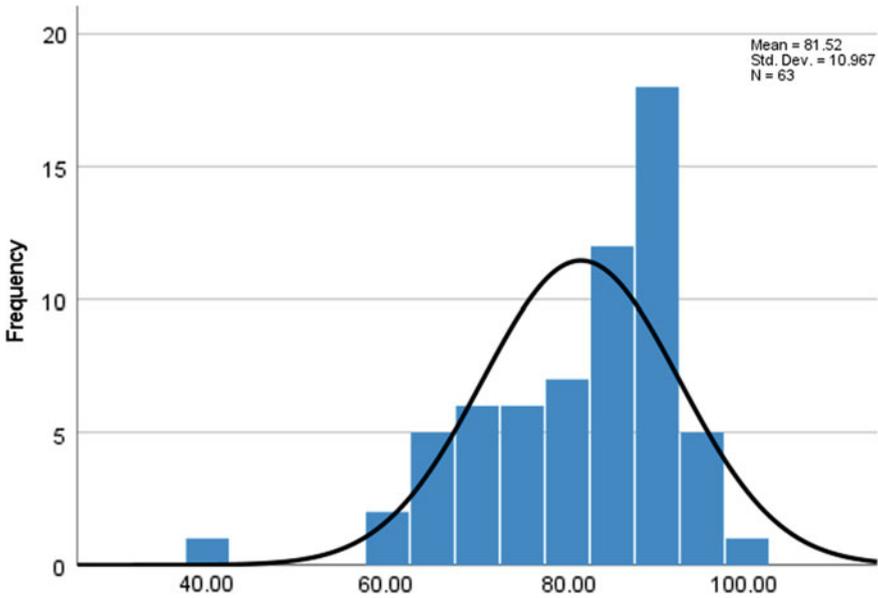Mean = 81.52
Std. Dev. = 10.967
N = 63

Fig. 14.4 The distribution of the mean of assessment scores in the final exam
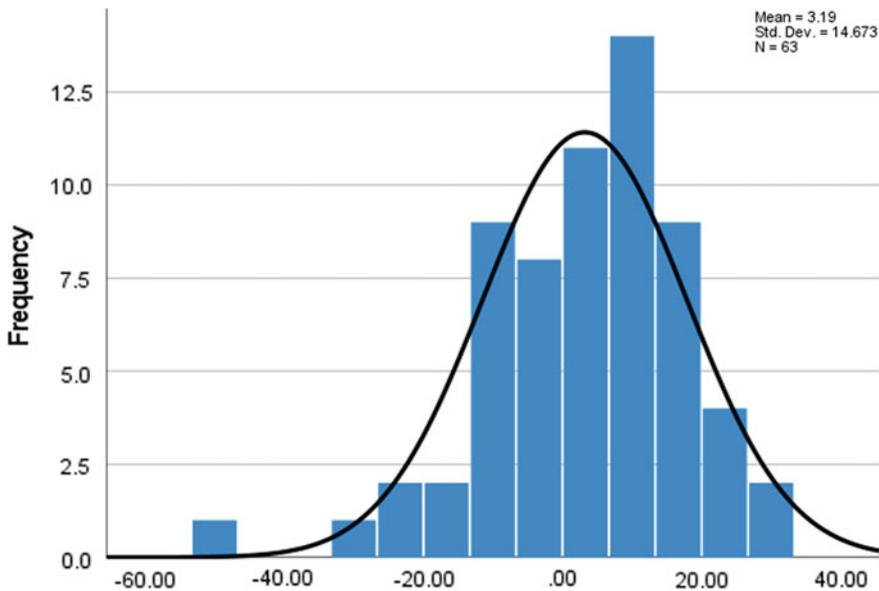
Fig. 14.5 The distribution of the difference between the two assessment scores for each student

Figure 14.5 shows the *distribution of the difference between the two assessment scores* for each student. The difference between the scores for each student is normally distributed (M = 3.19, SD 14.673, N = 63) which also satisfies assumption (iii).

With assumption (iii) satisfied, the remaining test is whether the difference between the two mean scores is statistically significant, and to what effect. Matched pairs t-test procedures were implemented on the IBM SPSS Statistics 28 to evaluate the impact of the assessment for learning instruction methods on students' final scores in the mathematical modelling course. Inferential procedures on the data revealed that there was no statistically significant increase in the assessment scores obtained during the AfL phase (M = 78.33, SD = 8.86) and the scores obtained from the assessment given at the end of the course (M = 81.52, SD = 10.97), $t(62) = 1.728$, $p = .089$, $\eta^2 = 0.218$. Using Cohen's (1988) conventions of 0.20, 0.50, and 0.80, respectively, for small, medium, and large effect sizes, the effect size of 0.22 corresponds to a small effect in practice.

14.4 Discussion and Conclusion

This chapter used a mixed pairs-study with pre-service secondary mathematics teachers as participants, to measure changes that would take place in their performance scores during the course and at the end, if assessment for learning, was used. Four assumptions A1-A4 above were tested using the quantitative data gathered

during the course (formative) and at the end of the course (summative). The study shows that, although the mean score for summative assessment is higher 81.52% (SD = 10.97, n = 63) than the mean score obtained during the formative assessment obtained during the AfL phase (M = 78.33, SD = 8.86), inferential statistical analysis revealed no significant increase in the assessment scores obtained during the AfL. This finding agrees with the study by Besser et al. (2013) who also found no significant difference between their control and experimental groups in their post-test data. However, Besser et al. reported that the control group performed significantly better in the pre-test than the experimental group, but in the post-test, the differences were no longer visible.

The findings in this pilot study proffer the idea that there is a difficulty in finding an assessment protocol in mathematics education in general, and mathematical modelling in particular, that is not only theoretically supported, but also effective in practice. While research output in mathematics education generally favours constructivist theoretical frameworks, often the assessments that would match such innovations are context-specific and difficult to replicate in other jurisdictions, leaving teachers in a dilemma.

A matched pairs design was adopted in this study to minimize the logistical requirements of splitting a one semester pre-service teachers modelling class into two and teaching them differently, one in the experimental and one in the control group. Instead, the same group was taught the same content, assessed at different times throughout the course with the overall aim of improving learning throughout the course up to its end. The findings revealed relatively higher mean scores both in the formative assessment and in the summative assessment, but the difference between the two mean assessments was not statistically significant. Also, the impact of the assessment for learning on the final score in terms of the effect size is small.

Was there a notable change in the pre-service teachers' mathematical modelling scores following a teaching plan that applies the assessment for learning (AfL) framework? Yes, the findings are encouraging in two main aspects. First, the AfL approach offers a very strong possibility of improving the students' gains during the learning sessions and also at the end of the sessions. Moreover, while the difference between the two assessment scores T_1 and T_2 was not statistically significant, both T_1 and T_2 showed relatively high mean scores suggesting good performance overall by pre-service teachers enrolled in the course. The high means in the two measurements can be considered as a contribution of AfL to the individual students' grades. The fact that the effect size was also found to be small is not surprising given that the difference between the two means was already not statistically significant. The current study contributes to research into assessment methods in pre-service mathematics education courses which include mathematical modelling courses and to understanding their practical contributions to the teachers' learning gains at the end of such courses.

References

Alagoz, C., & Ekici, C. (2020). Cognitive diagnostic modelling for mathematical modelling assessment. In G. A. Stillman, G. Kaiser, & C. E. Lampen (Eds.), *Mathematical modelling education and sense-making* (pp. 349–359). Springer.

Allal, L. (2010). Assessment and the regulation of learning. In E.B.P. Peterson (Ed.), *International encyclopaedia of education* (Vol. 3, pp. 348–352). Elsevier.

Antonius, S., Haines, C., Jensen, T. H., Niss, M., & (with Burkhardt, H.). (2007). Classroom activities and the teacher. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 295–308). Springer.

Andrade, H. L., & Heritage, M. (2018). *Using formative assessment to enhance learning, achievement, and academic self-regulation.* Routledge.

Berry, J. S., & Le Masurier, D. (1984). OU students do it by themselves. In J. S. Berry, D. N. Burghes, I. D. Huntley, D. J. G. James, & A. O. Moscardini (Eds.), *Teaching and applying mathematical modelling* (pp. 48–85). Ellis Horwood.

Besser, M., Blum, W., & Klimczak, M. (2013). Formative assessment in everyday teaching of mathematical modelling: Implications of written and oral feedback to competency-oriented tasks. In G. A. Stillman, G. Kaiser, W. Blum, & J. P. Brown (Eds.), *Teaching mathematical modelling: Connecting to research and practice* (pp. 469–478). Springer.

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161–177.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice, 5,* 7–73.

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do? In *The proceedings of the 12th international congress on mathematical education* (pp. 73–96). Springer, Cham.

Blum, W., & Leiß, D. (2007). How do students and teachers deal with modelling problems? In C. R. Haines, P. L. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical Modelling (ICTMA 12): Education, Engineering and Economics* (pp. 222–231). Horwood. https://doi.org/10.1533/978 0857099419.5.221

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Erlbaum.

Department of Basic Education (DBE), South Africa. (2011). *Curriculum and assessment policy statement: Mathematics grades 10–12.* Department of Basic Education.

Durandt, R., Blum, W., & Lindl, A. (2022). Fostering mathematical modelling competency of South African engineering students: Which influence does the teaching design have? *Educational Studies in Mathematics, 109*(2), 361–381.

Gan, Z., He, J., & Mu, K. (2019). Development and validation of the Assessment for Learning Experience Inventory (AFLEI) in Chinese higher education. *The Asia-Pacific Education Researcher, 28*(5), 371–385.

Geiger, V., Galbraith, P., Niss, M., & Delzoppo, C. (2022). Developing a task design and implementation framework for fostering mathematical modelling competencies. *Educational Studies in Mathematics, 109*(2), 313–336.

Greefrath, G., Kaiser, G., Blum, W., & Borromeo Ferri, R. (2013). Mathematisches modellieren—Eine Einführung in theoretische und didaktische Hintergründe. In R. Borromeo Ferri, G. Greefrath, & G. Kaiser (Eds.), *Mathematisches Modellieren für Schule und Hochschule* (pp. 11–37). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-01580-0_1

Greefrath, G., & Vorhölter, K. (2016). Teaching and learning mathematical modelling: Approaches and developments from German speaking countries. *Springer*. https://doi.org/10.1007/978-3-319-45004-9

Haines, C., & Dunthorne, S. (Eds.). (1996). *Mathematics learning and assessment: Sharing innovative practices.* Arnold.

Hall, G. (1984). The assessment of modelling projects. In J. S. Berry, D. N. Burghes, I. D. Huntley, D. J. G. James, & A. O. Moscardini (Eds.), *Teaching and applying mathematical modelling* (pp. 143–148). Ellis Horwood.

Hankeln, C., Adamek, C., & Greefrath, G. (2019). Assessing sub-competencies of mathematical modelling—Development of a new test instrument. In G. A. Stillman & J. P. Brown (Eds.), *Lines of inquiry in mathematical modelling research in education* (pp. 143–160). Springer.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hidayat, R., Adnan, M., Abdullah, M. F. N. L., & Safrudiannur (2022). A systematic literature review of measurement of mathematical modeling in mathematics education context. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(5), Article em2108. https://doi.org/10.29333/ejmste/12007

Houston, K. (2007). Assessing the "phases" of mathematical modelling. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 249–256). Springer.

Izard, J., Crouch, R., Haines, C., Houston, K., & Neill, N. (2003). Assessing the impact of teaching mathematical modelling: Some implications. In S. J. Lamon, W. A. Parker, & S. K. Houston (Eds.), *Mathematical modelling: A way of life ICTMA 11* (pp. 165–177). Horwood.

Kaiser, G. (2007). Modelling and modelling competencies in school. In C. R. Haines, P. L. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling (ICTMA 12): Education, Engineering and Economics* (pp. 110–119). Horwood. https://doi.org/10.1533/9780857099419.3.110

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice, 7*, 197–207.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice, 16*(3), 263–268. https://doi.org/10.1080/09695940903319646

Leong, R. K. E. (2012). Assessment of mathematical modelling. *The Journal of Mathematics Education at Teachers College, 3*(1), 61–65.

Lester, F. K., Jr. (2013). Thoughts about research on mathematical problem-solving instruction. *The Mathematics Enthusiast, 10*(1), 245–278.

McMillan, J. H. (Ed.). (2013). *Sage handbook of research on classroom assessment*. Sage.

Moore, D. S., Notz, W. I., & Fligner, M. A. (2013). *The basic practice of statistics* (6th ed.). Freeman.

Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and Applications in Mathematics Education. The 14th ICMI Study* (pp. 3–32). Springer. https://doi.org/10.1007/978-0-387-29822-1_1

OECD (2008). Assessment for learning: The case for formative assessment. [Paper presentation] *Learning in the 21st Century Research, Innovation and Policy*, OECD/CERI International Conference https://www.oecd.org/site/educeri21st/40600533.pdf

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Penrose, O. (1978). How can we teach mathematical modelling? *Journal of Mathematical Modelling for Teachers, 1*(2), 31–42.

Rakoczy, K., Klieme, E., Leiß, D., & Blum, W. (2017). Formative assessment in mathematics instruction: Theoretical considerations and empirical results of the Co2 CA project. In D. Leutner, J. Fleisher, J. Grünkom, & E. Klieme (Eds.), *Competence assessment in education: Research models and instruments* (pp. 447–467). Springer.

Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast, 10*(1), 9–34.

Stewart, J., Redlin, L., & Watson, S. (2015). *Precalculus: Mathematics for calculus*. Cengage Learning.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Wess, R., Klock, H., Siller, H.-S., & Greefrath, G. (2021). *Measuring professional competence for the teaching of mathematical modelling: A test instrument*. Springer. https://doi.org/10.1007/978-3-030-78071-5

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). Routledge.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3–14.

Yockey, R. D. (2011). *SPSS Demystified: A step-by-step guide to data analysis for SPSS Version 18.0*. Pearson Education.