BUEFAD | 2021, Volume 10, Issue 2, 293-308
Bartın University Journal of Faculty of Education
dergipark.org.tr/buefad
DOI: 10.14686/buefad.839841

Designing Web-Based "Measurement and Evaluation" Learning Modules for Teachers' Needs*

Selma Şenel[a**], Serpil Günaydın [b], & Bülent Pekdağ [c]

a* Dr., Balıkesir University (http://orcid.org/0000-0002-5803-0793), **selmasenel@balikesir.edu.tr

b Lecturer, Balıkesir University (http://orcid.org/0000-0001-6304-1107)

c Prof. Dr., Balıkesir University (http://orcid.org/0000-0003-0611-0617)

*This research was supported by Balıkesir University Scientific Research Projects Coordination Unit (2018/048).

Research Article | Received: 13.12.2020 | Revised: 29.01.2021 | Accepted: 07.02.2021

Abstract

The challenging aspects of scheduling face-to-face in-service training programs are a barrier to make needed training widespread and accessible. Due to the limitations, the necessity has arisen to design interactive and updatable digital learning materials for teachers that have been prepared by experts and that can be accessed whenever desired. This design and development research presents a design of digital learning materials that has been drawn up on the basis of scientific processes and in the knowledge of teachers' inadequacies as regards the topic of measurement and evaluation. The learning materials produced in this research, Web-based Measurement and Evaluation Learning Modules (W-MELM), are based on the critical aspects of item writing and test development and the needs of teachers based on the literature. The ADDIE model was taken as the basis of W-MELM in the design process. Five experts were consulted in leading to the revision of the modules to make them ready for operation. The evaluation stage consisted of using a Likert-type data collection tool of 18 items to collect teachers' (n=50) opinions regarding W-MELM. As a result of analysis, the teacher's views acknowledged that W-MELM that had been designed met their needs and was an effective learning tool.

Keywords: Measurement, evaluation, assessment literacy, ADDIE, teacher training, distance education

Öğretmen İhtiyaçlarına Dönük Web Tabanlı "Ölçme ve Değerlendirme" Eğitim Seti Tasarımı

Öz

Yüz yüze hizmet içi eğitimlerin maliyet, zaman ve program gibi zorlayıcı faktörleri, ihtiyaç duyulan eğitimlerin gerekli yaygınlığa ulaşılmasına engel teşkil etmektedir. Bu sınırlılıklar nedeniyle öğretmenlere alan uzmanları tarafından hazırlanmış, istedikleri zaman erişebilecekleri, tasarım ve yayın süreci dışında bir maliyet ya da iş yükü getirmeyecek, etkileşimli ve güncellenebilir dijital eğitim materyalleri tasarlanmasının gerekliliği ortaya çıkmaktadır. Bu araştırmada, ölçme ve değerlendirme konusunda öğretmen yetersizlikleri göz önünde bulundurularak ve bilimsel temeller ışığında, öğretmen eğitimine yönelik dijital bir eğitim setinin tasarımına ilişkin bir örnek sunulmuştur. Eğitim setinde, alan yazından yola çıkılarak, madde yazma ve test geliştirmedeki kritik noktalara ve öğretmenlerin gereksinimlerine odaklanılmıştır. Eğitim setinin tasarım sürecinde ADDIE (Analiz, Tasarım, Geliştirme, Uygulama ve Değerlendirme) modeli temel alınmıştır. Tasarım sürecinde; 3 ölçme ve değerlendirme, 2 bilgisayar ve öğretim teknolojileri eğitimi alanından olmak üzere toplam 5 uzman görüşü ve eğitim setinin kullanıcıları olan 3 öğretmenin görüşü alınarak eğitim setinde son düzenlemeler yapılmıştır. ADDIE modelinin değerlendirme basamağında dijital eğitim setine ilişkin öğretmen görüşleri (n=50), 5'li Likert tipinde 18 maddeden oluşan bir veri toplama aracıyla toplanmıştır. Araştırmada kullanılan veri toplama aracı eğitim setini, içerik, öğretim tasarımı ve değerlendirme olmak üzere üç temel boyutta değerlendiren maddelerden oluşmaktadır. Verilerin analizi sonucunda öğretmen görüşlerinin, genel olarak bu üç boyutun hepsi için olumlu yönde olduğu tespit edilmiştir. Bir başka ifade ile öğretmenlerin görüşleri tasarlanan dijital eğitim setinin, ihtiyaçlarına yönelik ve etkili olduğu yönündedir.

Anahtar kelimeler: Ölçme, değerlendirme, ölçme ve değerlendirme okuryazarlığı, ADDIE, öğretmen eğitimi, uzaktan eğitim

To cite this article in APA Style:

Şenel, S., Günaydın, S. & Pekdağ, B. (2021). Designing web-based "measurement and evaluation" learning modules for teachers' needs. *Bartın University Journal of Faculty of Education, 10* (2), 293-308. https://doi.org/10.1016/buefad.839841

1 | INTRODUCTION

Measurement and evaluation (ME) are the building blocks of the education system. ME are the resources with which preparedness is ascertained, training activities are shaped, and a determination is made as to the degree to which students have reached targeted learning outcomes. Invalid measurement results lead to faulty determinations about the level to which students have acquired their learning goals, and also contribute to erroneous decisions made as from the point of measurement. When a major element of the system of education becomes dysfunctional, the system cycle breaks down as a consequence.

The results of measurements may be used in decisions of vital importance (Popham, 2005; Thorndike & Tracy Thorndike, 2014). Effective measurement in the classroom is critical to the monitoring of learning, systematic progress and achievement (Marzano, 2006; Murchan et al., 2013). Teachers play key role in ensuring the validity of ME in the classroom. Therefore, standards are developed for teacher competence in ME (American Federation of Teachers, National Council on Measurement in Education, 1990), and teachers' assessment literacy has been frequently researched in the literature (DeLuca et al., 2016; Fulcher, 2012; Mertler, 2003; Mertler & Campbell, 2005; Plake et al., 1993; Quilter & Gallini, 2000; Xu & Brown, 2016). Standards have been recently transformed and formative assessment have been focused (Brookhart, 2011; Gotch & French, 2014). Considering the standards given in the literature and the topics discussed for the assessment literacy, the subjects that a teacher should have competence in ME can be listed as follows:

1. Identifying achievement criteria clearly
2. Choosing the appropriate ME tool
3. Developing and implementing item/tool (tasks, rubrics, etc.)
4. Reliable scoring and grading
5. Presenting effective feedback and interaction
6. Using assessment to increase student motivation and achievement

Teachers must therefore have adequate knowledge and experience in the subject of ME. It has been reported in studies, however, that teachers report they experience problems in various subjects and on different scales in ME, and that they believe themselves to be inadequate in this context (Benzer & Eldem, 2013; Çakan, 2004; Gaitas & Alves Martins, 2017; Mertler, 1999, 2009; Plake et al., 1993; Popham, 2005; Şenel, Pekdağ & Günaydın, 2018; Topkapı & Yılar, 2016). Inadequacies of teachers and real-time practice problems may arise in several topics of assessment literacy. The most critical one of these problematic topics is about choosing the appropriate ME method for the construct to be measured and developing/implementing item or measurement tool (Plake et al., 1993). Functionality of following topics is highly dependent on choosing appropriate ME method and developing items/tools.

Teachers prefer the practice of copying a measurement tool instead of devising a new one. Teachers believe that composing test items is too time-consuming so they will avoid to write items and tend to copy previously composed items (Popham, 2005; Şenel et al., 2018). Another matter is that teachers use ME to help students achieve in high stake tests (Popham, 2005). As a result, teachers use only certain types of test items. Since a large majority of high stake tests are composed of multiple-choice items, teachers are more likely to use multiple-choice items (Bayat & Şentürk, 2015; Güneyli & Abbasoğlu, 2015; Şenel, 2018a). Multiple-choice items however are limited in their capacity to measure higher-level cognitive skills. It is known that knowledge and skills at each level cannot be measured with all type of test items (Anderson & Krathwohl, 2010; Gültekin, 2014). Wrong selection and usage of item type prevents the determination of whether or not students have reached the targeted skills. In other words, measurement results in this case do not provide information as to the degree to which the instruction has met its goals and formative evaluations are no longer functional. Additionally, it is reported in the literature that frequent changes

made in measurement dimensions in educational programs (Hamurcu, 2018) lead to an increase in teachers’ perception of inadequacy (Benzer & Eldem, 2013; Şenel et al., 2018).

Pre-service teachers are offered ME based courses in teaching programs of education faculties. However, these courses are reported to be insufficient for real-life practices (Alkharusi et al., 2011; DeLuca & Bellara, 2013; Mertler, 2003; Wang et al., 2004). To eliminate the problems that teachers encounter in ME and to alleviate their perception of inadequacy, it is clear that there is a need for applicable and functional training programs (Plake et al., 1993; Şenel et al., 2018). Teachers who had in-service training about ME are found to have competence compared to those without training (Mertler, 2009; Plake et al., 1993; Xu & Brown, 2016). Educational institutions organize in-service training programs with the aim of improving teachers’ professional competence and understanding, ensuring unity in applications and instilling the knowledge, skills and behavior required by the advances made in the field of education (Ministry of National Education [MEB], 1995). In-service training programs are known to require a significant amount of material resources as well as specialized personnel. According to the 2017-2018 Statistics for Formal Education of the Ministry of National Education (MNE), the number of teachers in Turkey working under the auspices of the Ministry is 1,030,130 (MNE, 2018). This statistic indicates that a large number of in-service training programs designed to provide face-to-face training to all teachers would mean making a major allotment of specialized personal and material resources, which would constitute a heavy burden on the economy. Planning in-service training programs for teachers during the academic year tends to disrupt the flow of education. When specialists cannot lead the training sessions, it is difficult to reach the expected level of productivity. The literature points to the need for focusing on the fundamental needs of specific in-service training programs, urging that costly training programs should be avoided (Clark & Mayer, 2003).

Due to these limitations and challenges, it is believed that interactive and updatable digital training materials for teachers that have been prepared by experts in their fields, that can be accessed whenever desired, and that will not generate costs or workloads other than what is needed in the design and publication process will be useful and functional. Digital educational materials may be considered a distance education application. Teachers tend to prefer distance education programs due to their limited available time and lack of locational flexibility (Taşlıbeyaz, Karaman & Göktaş, 2014). At the same time, it is also known that teachers may need such easily accessible learning opportunities when they face specific challenges that cannot be generalized, such as working with special groups like students with special needs (Şenel, 2018b). Similarly, web-based systems that allow individual instruction and training are developed and used (Wang et al., 2004; Wang et al., 2008).

Considering the importance of ME, teachers’ inadequacies in ME and ineffective in-service training, developing educational materials for teachers is critical. Distance education and online tools are becoming more widespread throughout the world due to its ease of use and easy access (Ferdig et al., 2020). In the rapidly advancing world of technology today, it has become the duty of all educational institutions to take advantage of all the opportunities presented. It should be considered that devising effective learning materials for in-class measurement that all teachers can access at any time they wish would be a major enhancement to education. This study is aimed to develop an interactive web-based tool, present its design procedure and to identify its effectiveness by using it in in-service teacher training.

2 | METHOD

This study was carried out as Design and Development Research (DDR). Educational research is often inspired by theoretical frameworks, therefore they may have isolated from real problems of everyday life (Design-Based Research Collective, 2003). DDR, which was defined as systematic research into the processes of design, development and assessment (Richey & Klein, 2008), consists of two types of

research. The first type of DDR concerns focusing on the processes of designing and developing products and tools. The second type of DDR concerns the improvement, use and approval of models. The focus is on the validity and effectiveness of the technique that is being tested (Richey, Klein & Nelson; 2004). The present study is of the second type of DDR. Another field of DDR is focused on new technologies. These studies may help to create useful designs with the collaboration of engineering and method. (Wang & Hannafin, 2005). This study too concentrates on how developing technologies can be used the most beneficially in teacher training. For this, a set of digital learning materials, *Web-based Measurement* and Evaluation Training Modules (*W-MELM*) have been developed to eliminate teachers' deficiencies in ME. W-MELM are the training modules consists of 9 modules focusing on the subject of ME, which were developed in line with the aim of the research and in order to meet the needs of teachers in this respect. The strength of this material has been examined in terms of effectiveness and sustainability.

W-MELM is based on the core phases of the ADDIE (Analysis, Design, Development, Implementation and Evaluation) instructional design model and has been reported in this context. The ADDIE model is made up of five phases: analysis, design, development, implementation and evaluation (Çağıltay & Göktaş, 2013). An attempt was made to diversify the presentation and use of materials included in the developed digital learning material. It is known that diversifying presentation and materials in in-service distance education programs is an element that has an impact on motivation (Taşlıbeyaz, Karaman & Göktaş, 2014). This is because it is thought that learning materials appeal more to the senses and are expected to attract more attention in this way.

Research ethics

The data collection phase of this study were approved ethically in accordance with the decision taken at the meeting of Balıkesir University Social and Human Sciences Ethics Committee dated 11.02.2021 and numbered 2021/01.

3 | Findings: ADDIE Phases

Analysis

The learning needs of the target group should be identified in the analysis phase and the desired goals of the training should be determined, taking into consideration internal and external dynamics. A scan of the literature was first carried out in the study to identify which matters pertaining to ME most needed instruction and should be included in the training materials (Bayat & Şentürk, 2015; Benzer & Eldem, 2013; Çakan, 2004; Gaitas & Alves Martins, 2017; Güneyli & Abbasoğlu, 2015; Şenel, 2018a; Şenel et al., 2018; Topkapı & Yılar, 2016). Choosing the appropriate technic and developing/implementing test or measurement tool is pre-requisite for assessment literacy. The scores obtained by teachers who do not have competence in assessment literacy are not valid, and feedback/communication will not contribute to the quality of education. The preliminary studies conducted by the researchers indicating their realization of the need for the present study (Şenel et al., 2018) were an important resource in the process of analysis. In this preliminary work, the deficiencies and problems teachers faced in the matter of ME were collected under 11 subheadings. The issue that most frequently came up in the interviews were their deficiency in writing test items (21.86%). It could be seen that the teachers tended to copy items (18.03%) because the writing process was long and difficult. Since the subject that is most commonly addressed in the literature is the process of writing items and developing tests, "Types of Items and Test Layout" was taken as a basis for this design and development research. Based on the belief that any information presented on the use of types of items would not be complete without an explanation as to the way in which open-ended items would be scored, it was thought useful to include *rubric*. In the light of the points emphasized in the literature, the content headings to be considered in the development of W-MELM are the following:

1. General introduction to the types of items
2. True-false items

3. Matching items
4. Multiple-choice items
5. Open-ended items with short answers
6. Open-ended items with long answers
7. Rubric
8. Test Layout

Design

The teaching aims for the learning system to be developed are determined in the design phase and suitable content is selected and prepared while educational strategies are ascertained (Çağıltay & Göktaş, 2013). The output of the analysis phase consists of the aims of this stage. In the design phase, content was prepared that was in line with the target topics that had been determined. Two academic researchers, one an expert in ME, the other in educational technologies, drew up the content. In preparing the content, the direct quotes of teachers appearing in the literature were used as a resource. The content was lined up hierarchically using MS PowerPoint in such a way that each topic constituted one module. In order to keep motivation up, critical issues were addressed so that the content could be prepared in a short space of time. In the content design process, it was believed that it would be useful to draw up important information that was common to every type of item as a separate teaching tool. Because of this, a module called “Important Information” was added to the subject headings determined in the analysis phase. After the content of the digital learning materials became clear, the design phase was entered, the process of which can be summarized as the following steps:

Deciding on Visual Composition

At this point, the visual aspects of the training were considered. Color charts were scanned for this purpose. Care was given to use no more than four colors on the screen (Yalın, 2000, p.99) and a decision was made as to which main colors would be used. In the selection of colors, colors that would be motivating to both men and women were chosen. At the same time, attention was also paid to make sure that the supplementary documents used in some of the modules would be of the same color and visual fabric. These supplementary documents were prepared as downloadable files, which ensured that users would always be able to have these materials on hand to consult. This content was designed to be printable infographic material. Infographic material refers to the visual representation of information, making it easier for the user to use and understand the data. Visual design was first prepared in the form of an MS PowerPoint presentation, but later this data was transferred to the final product. Common pages were designed for each module for the purpose of standardizing and organizing the instruction.

Creating Commonly Designed Pages

All of the modules were provided with a common structure in order to ensure wholeness and interface consistency, so that users could more easily understand the pages. In this context, the introductory page to each module, the page where the aim was described, a maxim appropriate to the content, item sample pages, pages with direct quotes from teachers, Q&A pages and conclusion pages were all designed to appear in all modules.

Deciding on a Tool for Developing Content and Formatting

Research was carried out on what tools had been used in the development of digital learning material, paying attention to those that permitted the design of pages that could be freely browsed and of materials that could be used for different learning modules. It was decided that the Articulate 360 tool would be used since this e-learning software met all the requirements. Different designs suitable to the content were chosen and while 7 of the 9 modules were prepared on a page format, the remaining 2 were

developed as video content. The modules were set up on a player device. The color scheme of the player was set up to match the general visual design. A hierarchical content tree was drawn up and configured so that the user could move between the pages in any way desired. Figure 1 displays W-MELM's interface, tree-structured menu and the player.

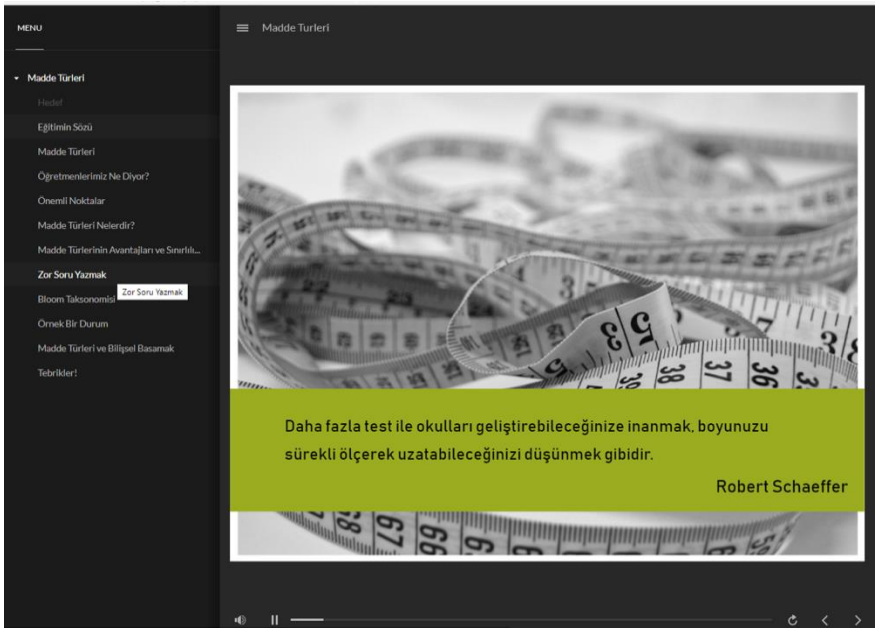


Figure 1. *W-MELM Interface: Tree-structured Menu and Player*

Development

In the development phase, the design drawn up in the previous stage is translated into an applicable product (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz & Demirel, 2014). When developing learning materials, a combination of different learning and teaching theories and approaches can be used to ensure that the instruction reaches its goal (Ally, 2004). In this study as well, benefit was drawn from Skinner's *Principles of Programmed Instruction* (Hergenhann, 1988), *Events of Instruction* (Gagne & Briggs, 1979) and *Multimedia Design Principles* (Clark & Mayer, 2003).

The first of the principles that is generally considered in the development process of digital learning materials is *Principles of Programmed Instruction*. According to Skinner (Hergenhann, 1988), the essentials of programmed instruction are to present knowledge in small increments, provide the individual with immediate feedback as to the accuracy or incorrectness of his/her learning, and allowing the individual to progress at his/her own pace. In W-MELM, each type of item was treated separately, in the question&answer screens, an answer was sought to a question on each page and gave the students immediate feedback. Additionally, the player in which W-MELM is embedded offers users a variety of features such as the ability to repeat material, go forward or backward, whichever is needed. This allows the individual to progress at his/her own pace. It can be said that W-MELM devised in this way satisfies the principles of programmed design.

Another set of principles used in the development phase is *Events of Instruction*. These principles stipulate that the learner must be informed of the goal (Gagne & Briggs, 1979). Users who know what they will be learning will prepare themselves both cognitively and affectively for the content and more easily participate in the learning process. This is why students are informed about the subject to be treated in the module at the beginning of each module.

Another theory that was of help in developing the digital learning material was *Multimedia Design Principles*. According to cognitive theory and the results of various studies, it is recommended that digital learning material is not composed solely of text but of a combination of text and visual elements (Clark &

Mayer, 2003, p.54). Priority was given in all of the phases of devising W-MELM to having a combination of text, visuals and graphical elements appear together (Figure 2).

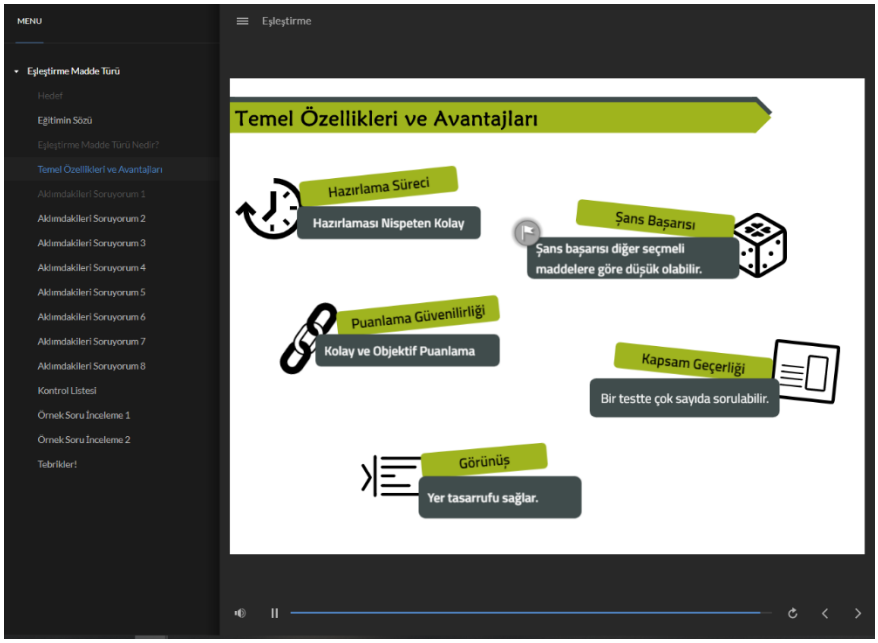


Figure 2. *The Use of Words and Graphics Together*

Another aspect of the principles of designing media is personalization. This principle is based on the knowledge that statements offered in the form of a dialog are more effective than narrative text (Clark & Mayer, 2003, p. 134). Taking this principle into consideration, the pages were designed in such a way that real-life problems were matched with a teacher’s profile (Figure 3). It is known that users can identify with this type of design and the real-life situation it presents, leading to effective learning. An effort was made on each page to provide teachers with answers to their questions about item types with as many examples of real-life situations as was possible. Making use of numerous examples in designing digital learning materials is known to enhance retention of what is learned (Clark & Mayer, 2003, p.175).

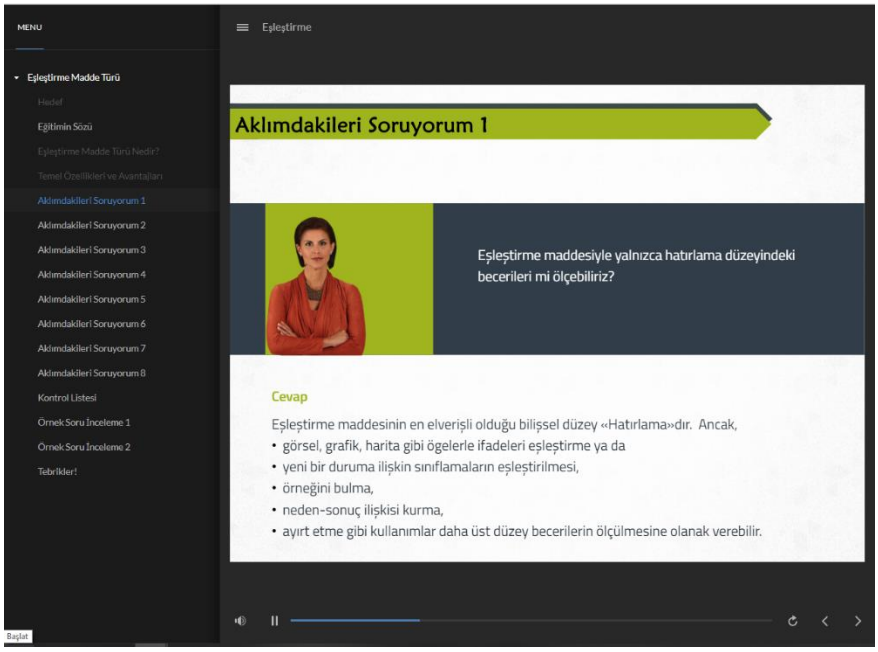


Figure 3. *Display of Quotes in the Form of Dialog*

In the sample situations pages, which were prepared under the theme of "I'm asking about what's on my mind," users are provided with the opportunity to seek answers to the questions in their minds about ME. It is believed that this ensures better learning of the type of item being discussed. Following these pages, users are directed to "Examining Sample Items" pages, where they are able to test their knowledge (Figure 4). Users on these pages are expected to correctly draw up a type of item and evaluate required items according to a checklist.

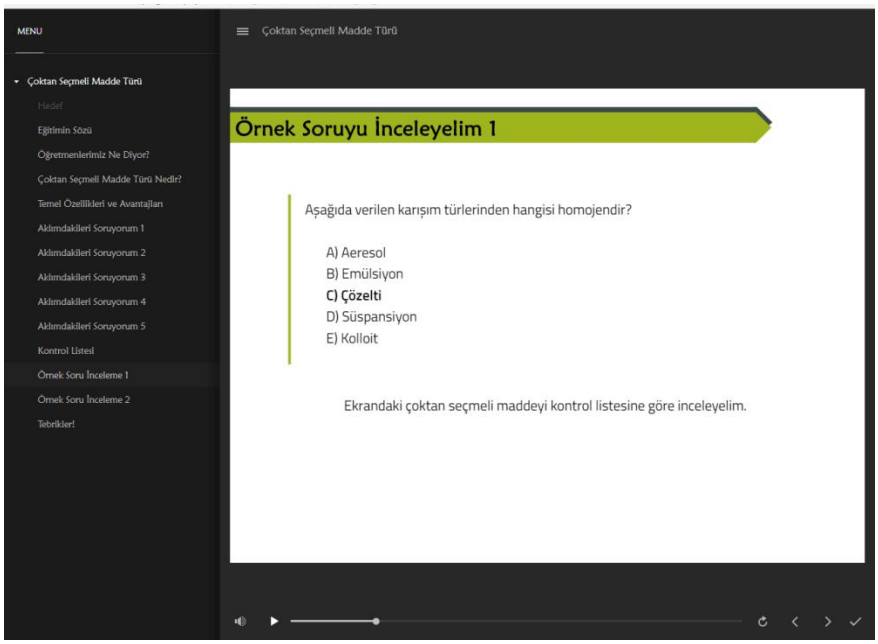


Figure 4. *Examining Sample Items*

Also added to these pages are useful downloadable documents on matters that teachers may have a need for and feel would be good to have on hand. Comparison tables or summarized information that teachers would want to keep at hand are presented here in the form of graphs (Figure 5). The design of these pages was set up on a template to preserve integrity.

Figure 5. *A Sample Downloadable Learning Document*

The testing format and critical information to learn is presented in W-MELM in the form of video content. This content was organized in the Camtasia 8 software, which facilitated sound and animation synchronization. In the development stage, five experts were consulted in the context of three measurement and evaluation, two computer and instructional technologies training fields, after which 3 teachers who were users of the training set were consulted for their opinions, leading to the revision of W-MELM modules to make them ready for operation.

Implementation

The teachers in the study group with whom W-MELM was implemented were selected with the convenience sampling technique. A total of 50 teachers in different branches were contacted. These teachers were from the branches of Chemistry (n=33), Mathematics (n=6), Science (n=3), Turkish or Turkish Language and Literature (n=3), English (n=2), Physics (n=2), and one (n=1) was a Homeroom Teacher. The teachers who were acting as raters in the study group included those who were new graduates as well as those with 35 years of teaching practice. The average seniority of the study group was 10 years ($\bar{X}$=9.96), and it can be said that they were a heterogeneous group in terms of their seniority (S=10.45).

Evaluation

An assessment form for evaluating the devised W-MELM was drawn up by the researchers who had been working in scientific studies on measurement and education technologies. In addition to the items intended to assess W-MELM, the form also included information on the seniority and branch status of the teachers. The items on the form focused on the points targeted in the development stage of W-MELM and on the needs of the teachers. The form consists of 18 items devised to understand whether or not W-MELM meets the aims of the research. The items were devised so as to focus on three fundamental dimensions. These are: (i) Whether or not W-MELM content (*Content: Items 1-5*), (ii) and the training design principles had been effectively implemented (*Instructional design: Items 6-12*), and (iii) a general assessment of W-MELM and whether it can be expanded for widespread use (Evaluation *and Conclusion: Items 13-18*). The responses are structured on a 5-point Likert-type scale (1-I definitely disagree, 2-I disagree, 3-I partly agree, 4-I agree, 5-I definitely disagree). An open-ended item was added to the end of the form for the assessor to provide any comments they may have about the study.

The teachers who were the last users of W-MELM were asked for their evaluation of the final product that emerged from the design development. The form was applied to the study group described in the implementation step. After the implementation of W-MELM, the teachers were asked to fill out the assessment form that had been created as part of the research. The Cronbach alpha coefficient of the items was found to be 0.93. Histogram chart for the total scores is presented in Figure 6.

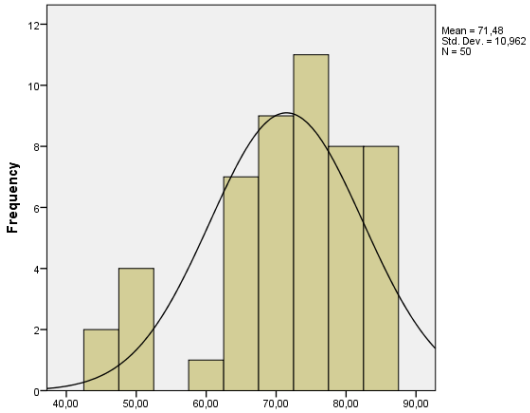
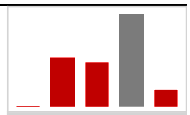
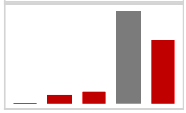
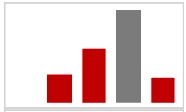
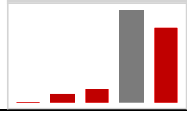


Figure 6. *Histogram Chart for the Distribution of Scores*

Descriptive statistics regarding the total scores obtained from the form were calculated. The highest score that can be obtained from the form is 90, the minimum score is 45, and the range is 42. The average of the total scores is 71.48 and standard deviation was found to be 10.96. Considering the the graph and the statistics obtained, it can be stated that the distribution is skewed to the left and tends towards positive opinions.

The $\bar{X}$ and S obtained from the responses to each item of the teachers are presented in Tables 1, 2 and 3, which display, in mini graphs, the frequencies and percentages for each response and their distribution by category. The first point that is striking in the review of the tables is that all of the items with the highest response frequency had been answered by "*I agree*" or "*I definitely agree*." Since all of the items in the form had been created as positive items of assessment regarding W-MELM, this outcome is an important indication that W-MELM had reached its targeted goal. Furthermore, the mean score for the responses to the items, varying in the range of 3.08-4.42, shows that the opinions of the large majority of the teachers varied in the range of "I partly agree (3) - I definitely agree (5)." Looking at standard deviation, it can be seen that outside of item 3, standard deviation is below 1. This reveals that the teachers' views did not deviate severely from the mean.

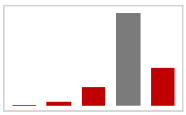
Table 1. Descriptive Statistics on the "Content" Dimension

No	Item	$\bar{X}$	S	I definitely disagree		I disagree		I partly agree		I agree		I definitely agree		Mini Graph of Distribution of Responses by Category (1-5)
				f	%	f	%	f	%	f	%	f	%	
1	I realized I made mistakes in writing up items.	3.30	1.07	2	4	11	22	10	20	19	38	5	10	
2	I learned the critical points to watch out for in writing up items.	4.06	0.93	1	2	3	6	4	8	23	46	16	32	
3	The checklists that reminded us of what features each type of item had to contain was important in the training.	4.42	0.78	0	0	1	2	5	10	14	28	27	54	
4	The questions on the "I'm asking about what's on my mind" pages were really the ones that had been on my mind too.	3.52	0.91	0	0	7	14	13	26	22	44	6	12	
5	The content addresses the goal of the instruction.	4.10	0.93	1	2	3	6	4	8	21	42	17	34	

Category with the highest frequency

Table 1 shows that 3 of the 5 items have a mean of over 4. The teachers stated that with W-MELM, they had learned the critical points to watch out for in writing up items ($\bar{X}_2$=4.06; S_2=0.93) and that the set addressed the goal of the training ($\bar{X}_5$=4.10; S_5=0.93). Moreover, it can also be said the teachers found the checklists to be helpful in underlining which features each type of item should have ($\bar{X}_3$=4.42; S_3=0.78). It was seen that the teachers selected the *I realized I made mistakes in writing up items* choice at a relatively lower rate ($\bar{X}_1$=3.30; S_1=1.07).

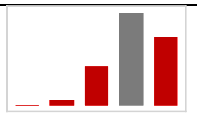
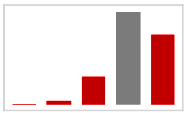
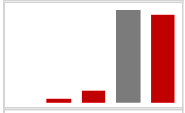
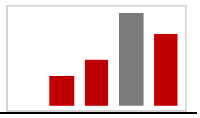
Table 2. Descriptive Statistics on the "Instructional Design" Dimension

No	Item	$\bar{X}$	S	I definitely disagree		I disagree		I partly agree		I agree		I definitely agree		Mini Graph of Distribution of Responses by Category (1-5)
				f	%	f	%	f	%	f	%	f	%	
6	The training was designed in an attractive way.	4.02	0.89	1	2	2	4	4	8	27	54	13	26	
7	The presentation of the content (sound/visuals/text) was effective.	3.96	0.90	1	2	2	4	6	12	26	52	11	22	
8	Stating in each module what we would be learning increased our motivation.	4.20	0.83	0	0	2	4	4	8	21	42	20	40	
9	Each module was tied to what was previously learned and to the other modules.	4.20	0.67	0	0	0	0	4	8	28	56	14	28	
10	The examples of items and the feedback reinforced my learning.	4.30	0.76	0	0	2	4	2	4	22	44	21	42	
11	The medium allowed me to practice what I had learned.	3.92	0.75	0	0	2	4	10	20	26	52	9	18	
12	It was effective to have the responses assessed and a result stated.	4.28	0.67	0	0	0	0	6	12	22	44	19	38	

Category with the highest frequency

Table 2 shows that all of the teachers' responses displayed mean scores of approximately 4 and over. This outcome indicates that W-MELM had served its educational design purpose. In particular, this suggests that the teachers' examples of items and the feedback were effective in the learning process ($\bar{X}_{10}$=4.30; S_{10}=0.76). Additionally, the teachers found it effective to have a construct where responses would be assessed in W-MELM ($\bar{X}_{12}$=4.28; S_{12}=0.67). The high mean scores for the responses to these two items may be interpreted to indicate that the teachers actually preferred to have interactive digital learning materials. In general, it can be said that the teachers found the instructional design dimension attractive and effective. Descriptive statistics of the "evaluation and conclusion" dimension of the assessment form is presented in Table 3.

Table 3. Descriptive Statistics of the "Evaluation and Conclusion" Dimension of the Assessment Form

No	Item	$\bar{X}$	S	I definitely disagree		I disagree		I partly agree		I agree		I definitely agree		Mini Graph of Distribution of Responses by Category (1-5)
				f	%	f	%	f	%	f	%	f	%	
13	I would prefer this digital instruction material to be used in in-service training.	3.96	0.97	1	2	2	4	9	18	20	40	15	30	
14	I would prefer other similar interactive instruction to be used in in-service training.	4.02	0.96	1	2	2	4	7	14	21	42	16	32	
15	This digital instruction material is more effective than face-to-face in-service training programs.	3.08	1.12	3	6	14	28	12	24	14	28	5	10	
16	Teachers should be offered online training in areas other than measurement.	4.30	0.76	0	0	1	2	3	6	22	44	21	42	
17	I would prefer in-service training to be in a format that we can always access.	4.00	0.90	1	2	2	4	5	10	27	54	13	26	
18	Since it is hard to be sure of the competence of the specialist involved in face-to-face in-service training programs, I would prefer online instruction that has been prepared by experts.	3.84	1.02	0	0	6	12	9	18	18	36	14	28	

■ Category with the highest frequency

It can be seen from Table 3 that teachers have positive opinions about making this and similar digital learning materials more widespread. It is observed that the teachers had the highest response rate on the question regarding providing teachers with online learning for other topics outside of ME ($\bar{X}_{16}$=4.30; S_{16}=0.76). At the same time, the findings show that the teachers would prefer similar in-service training programs that they could access whenever they want ($\bar{X}_{14}$=4,02; S_{14}=0,96; $\bar{X}_{17}$=4,00; S_{17}=0,90). The lowest mean score on the assessment form was in the item "*This digital learning material is more effective than face-to-face in-service training programs.*" ($\bar{X}_{15}$=3,08; S_{15}=1,12). This can be interpreted to mean that teachers also consider the advantages of face-to-face in-service training. The responses the teachers gave to the open-ended item on any more comments they would like to make clarify this point. One of the teachers in the study group said,

> This is a wonderful example of distance education for in-service training. But I also think that face-to-face education is necessary. That is, I think that the program can be produced on the basis of first offering the theoretical part of the topic to be learned in the form of face-to-face training and then the practice can start and continue for a certain period of time.

4 | Discussion & Conclusion

In this study, the deficiencies and needs of teachers in the topic of ME were considered and a scientific process was followed to design digital learning material. The dimensions of content, instructional design as well as that of evaluation and conclusion show that the teachers positively view on W-MELM and the

material satisfies the aims of the instruction. This result coincides with the literature reports positive outcomes of teachers who had ME training (Brookhart et al., 2010; Mertler, 2009; Plake et al., 1993; Xu & Brown, 2016). Additionally, results of this study draw parallel with research use web based materials to enhance assessment (Wang & Hannafin, 2005; Wang et al., 2008). Teachers who are able to choose the appropriate ME and develop/implement item or measurement tool (tasks, rubrics, etc) may be considered to have completed an important topic of assessment literacy (DeLuca et al., 2016; Fulcher, 2012; Mertler, 2003; Mertler & Campbell, 2005; Plake et al., 1993; Quilter & Gallini, 2000; Xu & Brown, 2016). Teachers can carry out the further steps in the light of valid results if the measurement practices are conducted thoroughly. Therefore, scoring, giving feedback and evaluation are latter competencies. Hereby, the skills that W-MELM focused on critical topic of teachers’ assessment literacy.

The validity and reliability of a measuring instrument is significantly dependent on whether or not the instrument contains the features that the items are required to have (Crocker & Algina, 1986; Gültekin, 2014). In this context, the checklists that were drawn up to provide guidance on whether or not the types of items in W-MELM had the required critical features received a high percentage of positive views. Also, seeing an actual listing of features of items, which is an essential factor in making valid and reliable measurements, proved to be an effective tool for the teachers.

It is known that it is important to make a needs analysis (Peterson, 2003), to follow scientific evidence and theories (Reigeluth, 1983), and to use technology effectively in the design process of digital learning material (Hooper & Rieber, 1995; Sezer, Karaoğlan Yılmaz & Yılmaz, 2013). W-MELM designed along these lines is believed to be a solution for ineffective and nonproductive face-to-face in-service training programs (Şenel et al., 2018; Uçar & İpek, 2006).

Literature emphasizes that teachers are demanding distance education now to a greater extent (Çelen, Çelik & Seferoğlu, 2013). Also, some challenging events as COVID-19 pandemic have made it necessary to conduct these trainings remotely (Ferdig et al., 2020). It is evident that distance education and qualified online materials will be needed more often in the future. The present study provides evidence that designing digital learning material is important and in fact required for today’s in-service training programs. Additionally, it can be seen that teachers actually prefer practical in-service training. Teachers not only want to be able to access effective digital learning materials whenever they wish to, they would also like to participate in face-to-face practice-oriented training. On the other hand, constantly accessible digital learning materials that were designed on the basis of scientific processes tend to be preferred because of the advantages of their being available at low cost, being based on a high level of expertise and with no time restrictions.

It was concluded in the study that a digital learning materials designed on the basis of scientific processes would be an effective tool in enhancing the qualifications of teachers in implementing classroom ME procedures. The study will provide guidance to researchers and implementers as to the processes involved in designing a digital learning material. At the same time, it might be recommended that such digital learning materials be expanded to be used in teacher training. There is a need for experimental research to explore the productivity and effectiveness of using digital learning materials in particular topics. In addition, the size of the working group in evaluation of W-MELM is limited. The effectiveness of W-MELM can be re-tested with larger groups. Reaching a wider audience may ensure the common use of the W-MELM. Views of the practitioners can be collected regularly to improve the software and content.

REFERENCES

Ally, M. (2004). Foundations of educational theory for online learning. *Theory and practice of online learning, 2, 15-44*.

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, *39*(2), 113–123. https://doi.org/10.1080/1359866X.2011.560649

American Federation of Teachers, National Council on Measurement in Education, N. E. A. (1990). *Standards for teacher competence in educational assessment of students*.

Anderson, L. W., & Krathwohl, D. R. (2010). *Bloom'un eğitim hedefleri ile ilgili sınıflamasının güncelleştirilmiş biçimi [A revision of Bloom's taxonomy of educational objectives]*, (Durmuş Ali Özçelik, trans.). Ankara: Pegem Academy Publishing.

Bayat, S., & Şentürk, Ş. (2015). Fizik, kimya, biyoloji ortaöğretim alan öğretmenlerinin alternatif ölçme değerlendirme tekniklerine ilişkin görüşleri [*Physics, Chemistry, Biology Teachers' Views on Alternative Assessment and Evaluation Techniques in Secondary School*]. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, *4*(1), 118-135.

Benzer, A., & Eldem, E. (2013). Türkçe ve edebiyat öğretmenlerinin ölçme ve değerlendirme araçları hakkında bilgi düzeyleri [Level of the Information About Turkish and Literature Teachers' Measurement and Assessment Materials]. *Kastamonu Eğitim Dergisi*, *21*(2), 649-664.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy and Practice*, *17*(1), 41–58. https://doi.org/10.1080/09695940903565545

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2014). *Bilimsel araştırma yöntemleri [Scientific research methods] (17th ed.)*. Ankara: Pegem Academy Publishing.

Clark, R. C., & Mayer, R. E. (2003). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. Pfeiffer.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim [Comparison of elementary and secondary school teachers in terms of their assessment practices and perceptions toward their qualification levels]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *37*(2), 99-114.

Çağıltay, K., & Göktaş, Y. (Eds.). (2013). *Öğretim teknolojilerinin temelleri: Teoriler, araştırmalar, eğilimler [Fundamentals of instructional technology: Theories, research, trends]*. Ankara: Pegem Academy Publishing.

Çelen, F. K., Çelik, A., & Seferoğlu, S. S. (2013). Analysis of teachers' approaches to distance education. *Procedia-Social and Behavioral Sciences*, *83*, 388-392.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education. *Journal of Teacher Education*, *64*(4), 356–372. https://doi.org/10.1177/0022487113488144

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8.

Ferdig, R. E., Baumgartner, E., Hartshorne, R., Kaplan-Rakowski, R., & Mouza, C. (2020). Teaching, technology, and teacher education during the covid-19 pandemic: Stories from the field. *Waynesville, NC, USA: Association for the Advancement of Computing in Education (AACE)*.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041

Gaitas, S., & Alves Martins, M. (2017). Teacher perceived difficulty in implementing differentiated instructional strategies in primary school. *International Journal of Inclusive Education*, *21*(5), 544-556.

Gagné, R. M., & Briggs, L. J. (1979). *Principles of instructional design.* Kruse, New York.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, *33*(2), 14–18. https://doi.org/10.1111/emip.12030

Gültekin, S. (2014). Testlerde kullanılacak madde türleri, hazırlama ilkeleri ve puanlaması [Test item, developing principles and scoring. N. Demirtaşlı (Ed.), *Eğitimde Ölçme ve Değerlendirme [Measurement and Evaluation in Education] içinde* (2nd ed.) (ss. 171-251). Ankara: Edge Academy.

Hamurcu, H. (2018). Comparative examination of the primary school science curricula in Turkey (Curricula of 1992, 2001, 2005, 2013 and 2017). *Journal of Education and Learning*, *7*(2), 261-279.

Herbenhahn, B. R. (1988). *Introduction to theories of learning. (3. Baskı).* New Jersey: Prentice Hall.

Hooper, S., & Rieber, L. P. (1995). Teaching with technology. In A. C. Ornstein (Ed.), *Teaching: Theory into practice* (pp. 154-170). Needham Heights, MA: Allyn and Bacon.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Marzano, R. J. (2006). *Classroom assessment & grading that work*. ASCD.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, *120*(2), 285. https://www.questia.com/read/1G1-59644154/assessing-student-performance-a-descriptive-study

Mertler, C. A. (2003, October). Preservice versus inservice teachers’ assessment literacy: Does classroom experience make a difference? *Annual Meeting of the Mid-Western Educational Research Association*.

Mertler, C. A. (2009). Teachers’ assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*(2), 101–113. https://doi.org/10.1177/1365480209105575

Mertler, C. A., & Campbell, C. (2005). Measuring teachers’ knowledge & application of classroom assessment concepts: development of the “Assessment Literacy Inventory.” *Annual Meeting of the American Educational Research Association*.

Milli Eğitim Bakanlığı. (1995). Milli eğitim bakanlığı hizmet içi eğitim yönetmeliği [Ministry of National Education in-service training regulation]. Resmi gazete: 8.04.1995/22252.

Milli Eğitim Bakanlığı. (2018). Millî eğitim istatistikleri, örgün eğitim 2017-2018 [National education statistics formal education 2017-2018]. Ankara: Turkish Statistical Institute.

Murchan, D., Gerry, S., Vula, E., Bajgora, A. G., & Balidemaj, V. (2013). *Formative assessment*.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice*, *12*(4), 10–12. https://doi.org/10.1111/j.1745-3992.1993.tb00548.x

Peterson, C. (2003). Bringing ADDIE to life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia*, *12*(3), 227-241.

Popham, W. J. (2005). Seeking redemption for our psychometric sins. *Educational Measurement: Issues and Practice*, *22*(1), 45–48. https://doi.org/10.1111/j.1745-3992.2003.tb00117.x

Quilter, S. M., & Gallini, J. K. (2000). Teachers’ assessment literacy and attitudes. *Teacher Educator*, *36*(2), 115–131. https://doi.org/10.1080/08878730009555257

Reigeluth, C. M. (1983). Instructional design: What is it and why is it. *Instructional-design theories and models: An overview of their current status*, *1*, 3-36.

Richey, R. C., & Klein, J. D. (2008). Research on design and development. In J.M. Spector, M.D. Merrill, J.van Merrienboer, & M.P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp.748-757). New York: Routledge

Richey, R. C., Klein, J. D., & Nelson, W. A. (2004). Developmental research: studies of instructional design and development. In D.H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 1099-1130). Manwah, NJ: Lawrence Erlbaun Associates.

Sezer, B., Karaoğlan Yılmaz, F. G., & Yılmaz, R. (2013). Integrating technology into classroom: the learner-centered instructional design. *International Journal on New Trends in Education & Their Implications*, *4*(4).134-144.

Şenel, S. (2018a). Programlama konusu "kazanımları" ile "sınıf içi ölçme süreçlerinin" bilişsel düzeylerinin karşılaştırılması [Comparison of programming "acquisitions" and "in-class measurement activities" in terms of cognitive level]. *International Journal of Computers in Education*, 1(2), 1-20.

Şenel, S. (2018b). *Öğretmenlerin kaynaştırma uygulamalarındaki ölçme ve değerlendirme yeterlikleri ve problemleri [Problems and insufficiencies of teachers' in educational measurement and evaluation of inclusive education]*. 6th International Congress on Measurement and Evaluation in Education and Psychology (September 5-8, 2018). Prizren, Kosova.

Şenel, S., Pekdağ B., & Günaydın, S. (2018). Kimya öğretmenlerinin ölçme ve değerlendirme sürecinde yaşadıkları problemler ve yetersizlikler [Chemistry teachers' problems and insufficiencies in educational measurement and evaluation]. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 12(1). 419-441.

Taşlıbeyaz, E., Karaman, S., & Göktaş, Y. (2014). Öğretmenlerin uzaktan hizmet içi eğitim deneyimlerinin incelenmesi [Examining the experiences of teachers received in service training through distance education]. *Ege Eğitim Dergisi*, *15*(1), 139-160.

Thorndike, R. M., & Tracy Thorndike, C. (2014). *Measurement and evaluation in psychology and education* (8th ed.). Pearson Education Limited. https://doi.org/10.2307/2282039

Topkaya, Y., & Yılar, B. (2016). Sosyal bilgiler öğretmenlerinin alternatif ölçme ve değerlendirme teknikleri hakkındaki görüşleri [Opinions of social studies teachers about alternative assessment and evaluation techniques]. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, *18*(1), 593-610.

Uçar, R., & İpek, C. (2006). İlköğretim okullarında görev yapan yönetici ve öğretmenlerin MEB hizmet içi eğitim uygulamalarına ilişkin görüşleri [Inservice education needs of class teachers teaching at elementary schools]. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *3*(1), 34-53.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23.

Wang, T. H., Wang, K. H., Wang, W. L., Huang, S. C., & Chen, S. Y. (2004). Web-based Assessment and Test Analyses (WATA) system: development and evaluation. *Journal of Computer Assisted Learning*, *20*(1), 59–71. https://doi.org/10.1111/j.1365-2729.2004.00066.x

Wang, T. H., Wang, K. H., Wang, W. L., Huang, S. C., & Chen, S. Y. (2008). Designing a Web-based assessment environment for improving pre-service teacher assessment literacy. *Computers and Education*, *51*(1), 448–462. https://doi.org/10.1016/j.compedu.2007.06.010

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Yalın, H. İ. (2000). *Öğretim teknolojileri ve materyal geliştirme [Instructional technologies and material development]*. Ankara: Nobel Academy Publishing.