Test-Based Blueprint Evaluation A Case Study on Software Engineering Course

Reem Alnanih

Faculty of Computing and Information Technology, Computer Science Department,
King Abdulaziz University, Jeddah, Saudi Arabi

Summary
The core concept of blueprinting has been widely employed to reproduce specific drawings in both academia and industry. In the modern electronic world, the concept of blueprinting has taken on a new shape. The intention behind creating a blueprint is to prepare a detailed plan of action before final execution. It is a systematic approach that ensures best educational practices. The purpose of this article is to validate the impact of applying a blueprint in exams—namely, in the final exam of a software engineering course in the Department of Computer Science at the Faculty of Computer and Information Technology in King Abdulaziz University, Saudi Arabia. This paper proposes a new and adaptable blueprint design that fulfils the requirements of both the course and the department. A controlled experiment was carried out among students of the Department of Computer Science in Fall 2017 and Fall 2018, respectively. The students were divided into two groups: The control group comprised the students who registered for Course Program Computer Science CPCS 351 in Fall 2017 and sat the final exam without using the blueprint method. The experimental group comprised the students who registered for CPCS 351 in Fall 2018 and sat the final exam using the blueprint method. An analysis of variance and a t-test were applied to compare the results of both groups' final exams. The results confirm that the use of blueprints in the final exam improved the performance of the students in the experimental group in comparison to the control group.

Key words:
Adaptable design, Blueprint, Content validity, Course learning outcome, Domain knowledge, Experiment test, Internal learning outcome, Software engineering, Test-based evaluation.

1. Introduction

In this era of computer-aided design, blueprinting remains one of the oldest techniques still in practice. Blueprinting, invented by John Herschel in 1842, refers to the process of contact printing on light-sensitive sheets to reproduce a technical drawing [1]. The term "blueprint" originates from the blue lines that appear on a white background as a result of the blueprinting process. It is a well-established technique that ensures the fast and accurate production of an unlimited number of copies. The core concept of blueprinting has been widely employed to reproduce specific drawings in both academia and industry.
Since its origins, blueprinting has been viewed as an easier, faster, and cheaper method of reproducing texts and drawings when compared to hand-tracing original documents. It costs one-tenth of the price of hand-traced reproductions, and it also eliminates the expense of photolithographic reproduction [2]. It was initially employed in the fields of construction and industry. Later, its use extended to car modeling and animation. In this era of newer and faster technologies, blueprinting is still used to create special artistic effects on fabric and paper.
As technology flourished, carbon copying was used in place of blueprinting for smaller documents. However, engineers, architects, and shipwrights continued to use blueprinting to reproduce large-scale drawings. Later, large-format xerographic photocopiers and diazo whiteprint processes replaced blueprinting to a large extent. Furthermore, due to modern printing methods, architects and engineers usually print their drawings from computers rather than undergoing the chemical process that makes the prints blue. Thus, today, blueprints are no longer "blue." As a result, they are more commonly referred to as drawings or plans.
Blueprinting exists in various professional fields. For instance, for circuit-level design in engineering, blueprinting offers the designer the freedom to choose the main parameters of the device used for a particular design (i.e., there are no preset procedural steps). Pinto and Maloberti [3] explored the transistor-level design of electronic circuits and employed a blueprinting tool in Java to provide the precise value of device parameters. The tool performed linear interpolation between the data points in their available ranges and generated an error message if a certain value appeared out of range. The precision in values depended on the resolution of the points. In the field of information security, Najjar [4] presented a low-cost, efficient, and practical blueprint for the implementation of public key infrastructure. The blueprint resulted in improved management and security, ensured compliance and long-term file retention, and reduced the amount of paperwork. In the modern electronic world, the concept of blueprinting has taken on a new shape. Generally speaking, the intention behind creating a blueprint is to prepare a detailed plan of action before final execution [5].
In the process of educational assessment, Newble et al. [6] defined a blueprint as a table specifying the key

Manuscript received November 5, 2019
Manuscript revised November 20, 2019

elements of performance evaluation such that the appropriate samples and corresponding methods of a particular activity can be selected accordingly. Blueprinting is also used by teachers who aim to provide an interactive approach to the achievement of learning objectives and curriculum expectations. In terms of career development and learning opportunities, course coordinators and evaluators also use blueprinting for curriculum development and for the design of learning plans. For researchers, it provides a clear framework through which to assess important concepts.

In the context of educational procedures, assessment systems are widely used. However, due to the evolution of network infrastructures and the widespread use of personal computers in recent years, the use of the associated technologies has had a significant impact on learning and assessment processes. Various assessment procedures and software tools have been developed for test-based evaluations of students' competencies, established through intended learning objectives (ILOs). Based on Bloom's taxonomy, a learning objective represents the desired knowledge level, understanding level, and application level in the students being taught. Furthermore, various technological solutions have been used to ensure assessment effectiveness [7]–[8].

Nowadays, outcome-based educational models are used in the National Qualification Framework (NQF) of Saudi Arabia and have been adopted by all Saudi universities. The NQF is a document developed by the National Commission for Academic Accreditation and Assessment (NCAAA), which is an autonomous body responsible for academic accreditation. The NCAAA defines the expected capabilities of graduates as well as the comprehensive learning outcomes of each education level [9]. The commission classifies the types of learning expected from students into five domains and describes the learning outcomes for each domain. These five domains are knowledge; cognitive skills; interpersonal skills and responsibility; communication, information technology (IT), and numerical skills; and psychomotor skills. These domains are described below [9]–[10]:

- Knowledge refers to the ability to remember, understand, and present information. It includes knowledge of specific tasks, concepts, theories, and principles.
- Cognitive skills refer to an individual's ability to apply their theoretical understanding of concepts and principles to the performance of various mental activities, including thinking, reading, and learning. Cognitive skills play a vital role in investigating the issues within a certain field of study and drawing valid conclusions based on these investigations.
- Interpersonal skills and responsibility are the behaviors individuals use to ensure effective interaction with others. These skills consist of the individual's intrinsic personality along with their continuous personal and professional development. Strong interpersonal skills can support an individual in numerous situations (e.g., during job interviews and in teamwork tasks).
- Communication, IT, and numerical skills are required for the effective transfer of information to others via verbal or written means. It involves the use of information and communication technology as well as mathematical techniques.
- Psychomotor skills are movement tasks that aim to enable an individual to learn about their environment. This fifth domain of learning applies only to certain programs and varies widely in its application. Because of its specificity, its learning outcomes have not been defined for each level of the NQF.

The proposed assessment method used in this article focuses on the first four domains classified by the NCAAA. While planning programs and performing assessments of students' learning, the following important points should be considered:

1. The desired profession or field of study of the students has a direct relationship with the learning outcomes of the first two domains defined above. General guidelines in terms of knowledge and expected skills are available in the framework for any field. However, program-specific knowledge and thinking skills must be identified when planning a particular program. The expected learning outcomes include evaluations concerning the student's field of study, a reliability test, and valid concluding remarks. When psychomotor skills are involved, the programs must specify the skills required for graduates.
2. In addition to program-specific expertise, students should develop the general capabilities described in the third and fourth domains above, irrespective of their corresponding fields of study. Special training courses can be organized to develop these capabilities, or they can be integrated into various courses. However, the capabilities developed through specially designed courses should also be extended to other studies. As a consequence, the development of these abilities is normally incorporated into teaching and learning processes and assessment criteria for all subjects.
3. The programs specific to the fourth domain of learning demand a high level of achievement. Such dedicated programs include those in the areas of languages, statistics, and mathematics.
4. A critical feature of the intended outcomes of each domain is that graduates will utilize the acquired

capabilities to conduct the described tasks and adopt them habitually in their personal and professional lives. Thus, these outcomes have significant implications for teaching strategies, student assessment, and program evaluation.

One of the key components of the above context is content validity, which ensures that a representative sample of ILOs is measured and that the candidates fulfil the minimum performance level (MPL) by reaching the intended competence level established in the ILOs. For a successful content validity test, each item of the assessment method must represent a minimum of one ILO. This test considers the curricular content and expected abilities of students and accordingly measures the assessment method's usefulness. Any systematic error in evaluation is an indication that either the course objectives or course delivery are not justified in terms of their content [11]–[12].

Exam blueprinting is a secure way of obtaining valid assessment results. It allows a department or institution to use a concrete examination strategy. Furthermore, for an assessment to be valid, presenting a careful combination of representative items is more important than building up high-quality items that represent ILOs. Thus, keeping in mind representativeness and combination issues, blueprinting is a systematic approach that ensures best educational practices [13].

The remainder of this paper is organized as follows. In Section 2, the educational goals are summarized and the motivations behind the design of the blueprint for assessment are discussed. Section 3 presents a detailed literature review to highlight the relevant research contributions in the domain of designing blueprints for assessment. Section 4 explains the proposed blueprint model. Section 5 presents the structure of designing adaptable blueprint. Section 6 validate the proposed model. Section 7 discusses the results, and Section 9 concludes the paper.

2. Motivations and Objectives

The purpose of this paper is to propose a method for designing blueprint assessments that are adaptable to the Department of Computer Science in the Faculty of Computer and Information Technology (FCIT) of King Abdulaziz University (KAU) in Jeddah, Saudi Arabia. The primary objectives of the proposed blueprint assessment design are to

- provide a conceptual map of the examination format and content area;
- ensure a speedy evaluation process to reduce the significant burden on academic staff;
- improve the fairness and reliability of the assessments;
- ensure the use of the same blueprint by course coordinators, instructors, and evaluators to minimize the possibilities of a "hidden curriculum"; and

ensure that content validity monitoring is performed through student feedback to guarantee teaching and exam reliability.

3. Literature Review

Assessment is mostly carried out through written examinations. However, this assessment method suffers from an unfair distribution of questions over topics, limited sampling, and vague questions, thus raising concerns about its validity. Blueprinting is a tool acknowledged as being capable of addressing these challenges and achieving the best possible educational practices. In this context, some highlighted research contributions are discussed below.

In [11], McLaughlin, Coderre, Woloschuk, and Mandin investigated the effect of publishing evaluation blueprints on students' perspectives of the evaluation process. A questionnaire consisting of items related to evaluation was developed to assess students' attitudes toward blueprints. Additionally, the data containing the overall course ratings, student performance, and the MPL for the evaluations were collected. A significant number of students acknowledged the course evaluation to be fair and reflective of the subject matter and delivered curriculum.

Ahmad and Hamed [14] created an exam blueprint and studied the impact of its utilization when evaluating students' achievement of learning objectives. The authors compared two groups of students: One group utilized blueprinting, while the other group did not. The responses collected through questionnaires revealed that the group of students who employed the proposed blueprinting method outperformed the other group. They achieved higher scores, showed higher satisfaction levels, and achieved their learning outcomes. The modified version of Kirkpatrick's model was used to evaluate the proposed method.

Patil, Gosavi, Bannur, and Ratnakar [15] proposed a blueprinting technique as a map for an assessment program to ensure that all aspects of the curriculum were covered by assessment programs over an indicated period of time. The proposed technique assigned appropriate weightages to all topics in the written examinations to ensure the content validity of the assessment. In order to prepare a blueprint, the purpose and scope of the assessment were defined, and then the weightage was assigned to the content areas. The calculation of the weightage was based on the perceived importance of the topic and its frequency of occurrence. The exam specifications were then decided, and the blueprint was prepared accordingly.

In [16], Becker and Vassar developed an assessment blueprint at the chapter level by defining the low-order and high-order cognitive objectives. The former corresponded to the recall level, while the latter corresponded to a combination of interpretation and problem-solving levels. For each chapter of the textbook, these two levels of cognitive objectives were associated with the proportion of time given to them in the classroom and at home. Based on these proportions, a test structure of 50 questions was prepared to calculate the number of desired questions based on chapter and content type. The proposed method was computationally straightforward and could be easily adopted by any instructor with a basic understanding of cognitive constructs. Furthermore, it was suggested that item writers should focus on central themes to prepare questions when developing a chapter-level blueprint.

In [17], Abdellatif and Al-Shahrani proposed a new technique to create a test blueprint based on total course-credit hours. The impact of the proposed method on the item analysis results was investigated in relation to the students' performance. For construct and content validity tests, the new method was compared with an existing method, where the designed test blueprint was based on the overall assessment time and the time permitted to answer each type of question. Both methods resulted in similar performances in terms of validity and reliability. Thus, the weight of topics based on actual contact hours, total assessment time, time for each item type, distinctive domain levels, and learning objectives were all declared key factors for a valid and reliable test blueprint.

There is an expanding body of literature that has reported on the importance of measuring the quality assessment of blueprints in terms of validity and reliability in the medical field. In [18], the authors conducted a study in a pharmacology department at a medical college in Gujurat, India. They prepared an assessment blueprint based on the internal syllabus. A total of 12 faculty members prepared learning objectives that were scored according to their clinical importance, and marks were distributed according to their relative weighting. The results of the study confirmed that blueprinting is an important aspect of written assessments in pharmacology education. In [19], the authors presented the feedback of medical teachers in relation to blueprinting. A large proportion of teachers were unfamiliar with the concept, and it was found that most written university exams do not follow the blueprint concept, despite its importance. In [20], the authors studied students at the College of Medicine in the University of Bisha, Saudi Arabia, to understand their perceptions of the blueprint as an assessment method in exams. The authors found that blueprints are an important aspect of outcome-based education. In [21], the authors proposed a guideline for designing and applying blueprints in the medical field. In [22] the author highlighted the importance of using blueprints when calculating the number of questions needed to reflect the weight of every topic. In [23], the extent to which English teachers in a Saudi school followed language-testing guidelines when developing blueprint tables for their formative and summative language tests was investigated. The results confirmed that the teachers usually prepare their exams without prior planning and rarely follow exam recommendations and specifications.

In [24] Gaffas et al., investigated the assessment method for the postgraduate training program in psychiatry in Saudi Arabia. Written exams of a set of MCQs were used as a summative assessment of residents to determine their eligibility for promotion from one year to the next. The authors developed a test based blueprint for preparing the set of questions for the exam. The authors confirmed that test blueprinting is an essential step to certify the test validity in all residency programs.

From the above research, it is clear that considering blueprints when designing and developing tests is important. It is also clear that there is a need to consider blueprints in the Department of Computer Science in the FCIT of KAU. None of the studies presented above have adapted blueprints to computer science courses, and neither have they defined the weight, domain, or time needed for each type of question in these courses.

4. The Proposed Model

In an educational system, a blueprint is a matrix (or matrices) that defines the relationship between an assessment item or question, its objective, and the appropriate tool for measurement. The questions in the assessment item are based on the ILOs from each topic in the content area. The matrix also identifies the percentage weighting of the cognitive dimensions.

In this paper, a straightforward and efficient process of designing a course blueprint has been proposed for the Department of Computer Science in the FCIT of KAU. The primary objective of the design is to validate the content of each assessment method to the defined articulation metrics of each course and guide the selection of learning experiences. An exam blueprint is already in use at the medical school at KAU. It is used as an assessment of test specification and evaluates time management. This means that assessments are conducted based on a reliable plan that is aligned to the structure of the blueprint in terms of course ILOs, domain knowledge, and assessment method. An adaptable blueprint is designed by linking the assessment method to the ILOs by preventing the under-representation of an ILO and avoiding assessments that are irrelevant and

inappropriate to the ILOs. The course learning outcomes (CLOs) are the ILOs that focus on what the students will learn from the course rather than what the teacher will teach. ILOs comprise a set of statements that specify what information students will know and be able to perform at the end of the specified course.
The blueprint approach aims to adapt the design of teaching to focus on what the instructor wants students to learn. It is an adaptable design approach to curriculum development that effectively satisfies certain requirements (Fig. 1) for the following reasons:

- It presents ILOs that explicitly state what the student is expected to be able to know, understand, or do.
- It presents learning activities that will help the student reach the ILOs.
- It assesses the extent to which the student meets these outcomes through the use of specific assessment criteria.

In this paper, the ILOs of the proposed design are based on the classification of knowledge, cognitive skills, and interpersonal skills [10]. The exam is based on two types of questions—namely, multiple choice questions (MCQs) and essay questions. MCQs are advantageous because they can test multiple knowledge areas in one question [25]. In some cases, we chose MCQs in the blueprint to indicate true/false responses. This method is also useful for questions with clear yes/no answers. Essay questions can assess a range of knowledge, skills, and cognitive levels. They can also assess the development of students' writing skills, as they allow students to display more knowledge on a selected point. However, these questions require more test time than MCQs.
In order to prove the concept of applying the blueprint in the assessment methods, the course considered for this study is CPCS 351, which aims to provide students with an understanding of the essential concepts of software engineering as they relate to system analysts. The course also guides students to design medium-scale software systems and learn how to apply engineering concepts and principles to real-world problems.

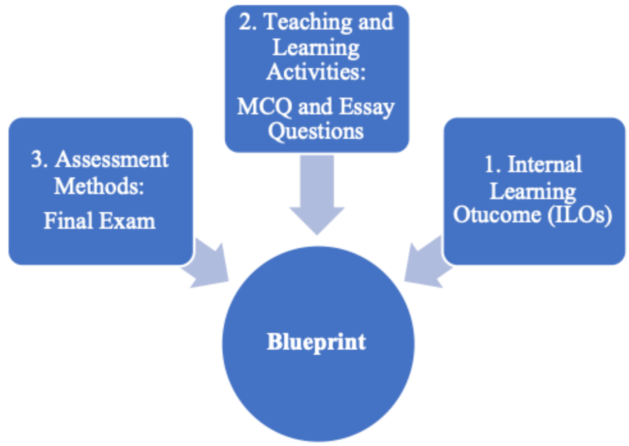


Fig. 1 Blueprint approach-based exam.

The steps involved in the proposed blueprinting design are explained as follows:

1) Content Analysis: A blueprint template is used, which consists of a sequence of rows and columns that define and systematically tabulate the curricular content and total number of evaluation items. A variety of tasks can be evaluated here, such as assignments, quizzes, and exams. These tasks highly depend on the nature of the course, thus meaning that they vary from course to course. However, they should be consistent with the ILOs of the relevant course. In the proposed design, the number of rows and columns are A and B, respectively (Table 1). The rows contain information about CLOs, whereas the columns define the attributes of blueprint.
2) Weight Assignment: All types of evaluations consist of a finite number of items; therefore, relative weights must be assigned to each content area so that the most important areas can be prioritized. An important concern in this step is to ensure the reliability of the weighting of the content areas. Weighting can be established through consensus, by taking inputs from other instructors, and through feedback from previous learners. In this paper, the weighting method is explained through letters E and G (Table 1).
3) Determining Question Types: After preparing the blueprint for content validity, the next task is to ensure that all types of evaluations, such as summative, objective, and retake evaluations, should conform to the blueprint. Creating new items for each evaluation type is a cumbersome task that involves extensive time and effort. However, it is often just a once-off investment that can pay off in the long run, as these items can be applied to other courses and shared with other departments.

4) Quality Assurance: In this step, evaluating actions are performed to ensure the high quality of the evaluation process based on the blueprinting weighting. This may involve the revision of some learning objectives to keep the hours, number of objectives, and number of evaluation items in proportion.

5. The Structure of Adapted Blueprinting

This section introduces the structure of our adapted blueprint. It also proposes the course structure in terms of ILOs. The proposed blueprint emphasizes the flexibility permitted in adapting curricula to institutional needs and according to the continual evolution of the field.

The structure of the proposed blueprint is illustrated in Fig.2 in the form of two entries (manual and automatic) that each have a set of letters, presented as follows:

- Manual entry (A, B, C, D, F, J, K). This entry depends on the coordinator of the course or the instructor who designs the exam and evaluates the students accordingly.
- Automatically generated entry (E, G, H, I). This entry is calculated based on the formal displayed below.

The guidelines for designing an adaptable blueprint are described as follows:

A. Include the following general information:

- course name
- course number
- exam type (e.g., quiz, final exam)
- date of exam
- total time allocated for MCQ exam (in minutes)
- total time allocated for essay exam (in minutes)
- total marks

A					E		F		G		H		I		J		K
B			C	D	MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	
Topic1	ILO1	What is Software?	CLO1	K	2	3	0	1	0	0.06	0	5	0	2	0	2	Q1.1
	ILO2	SW Methods		C	3	6	3	0	0.14	0	12	0	4.05	0	4	0	Q1.2
Total									1	1							

Legend: MCQ type | Essay type | Generated automatically | Entered manual | Exam paper

Fig. 2 Structure of the proposed blueprint

B. Include appropriate topics and ensure that

- the topics are organized according to the number of weeks in each term. For example, if the term consists of 14 weeks, the topics should cover each week (e.g., Topic 1, Topic 2, Topic 14);
- each topic is divided into a set of ILOs; and
- each ILO covers part of the main topic.

C. Include the CLOs and ensure that each CLO consists of a set of ILOs that is mapped to the relevant topic. The CLOs provide guidance on the depth of coverage of each topic.

D. Include the following learning domains (based on the NCAAA) [9]-[10] and can be applied to Computer Science department:

- knowledge
- cognitive skills
- interpersonal skills
- information technology IT and numerical skills

E. Define the type of questions used (whether MCQ or essay questions). Table 1 justifies the time needed (in minutes) to accomplish each type of question related to the domain type. It should be noted that the estimated time is based on an expert user who is experienced in the field of computer science and that the CPCS 351 course coordinator has been teaching the course for more than three years.

Table 1. Distributed time-based domain level

Domain level	Time taken (in minutes) to accomplish the specific domain level	
	MCQ	Essay
Knowledge (K)	2 m	3 m
Cognitive skills (CP)	3 m	5 m
Interpersonal skills (IP)	5 m	10 m
IT and numerical skills (IT)	5 m	10 m

F. Determine the importance level (IL). The IL scores each ILO item that has been tested by the MCQ/essay. The score can be one of the following values:

- Most important (the value of 3).
- Moderately important (the value of 2).
- Least important (the value of 1).
- Not included in the exam (the value of 0).

The total importance level (TIL)for the MCQs is the summation of all the values of all ILOs in all topics. The TIL for the essay questions is the summation of all the values for all ILOs in all topics.

G. Calculate the weight level (WL). The WL is the IL of each ILO/TIL, and the total weight level for each type of exam should be equal to one.

H. Calculate the duration of the domain. The duration depends on the domain type of each ILO and the exam type. The duration of the domain equals the WL multiplied by the total time (in minutes) allocated for the exam MCQ/essay.

I. Determine the number of items in each exam set. Here, one must illustrate how many questions are in the exam set of the assigned ILO.

- This is calculated as the duration of the domain in each exam set divided by the time (in minutes) taken to accomplish the specific domain level (see E).
- The number of items in each exam set is represented as a decimal.

J. Adjust for the number of items in each exam set:

- The number of items in each exam is adjusted by the instructor by rounding the decimal number to the nearest integer.

K. Map the ILOs to the exam paper. This column divides the ILOs and maps them to the exam paper by adding the question number to ensure that all the questions are distributed visually in the blueprint exam.

6. Validation Blueprint

Once the blueprint was ready for formal evaluation, it was necessary to conduct a pilot test so that it could be evaluated and any final changes to the proposed blueprint could be made. The pilot test was conducted with an expert user—a professor who has been working at KAU and using blueprints for more than eight years. The test was conducted in the professor’s office at KAU during normal working hours. In addition, an appointment with the professor was requested before the meeting.

The feedback from the professor confirmed to the researcher that designing and creating a new blueprint that is adapted to course ILOs and faculty requirements is a significant achievement. The professor recommended that the researcher apply the blueprint (Fig. 3). The domain times were also reviewed according to the professor’s expertise. A formal evaluation is required to validate the effectiveness of the blueprint on the designing exam. The following sub sections describe the formal evaluation.

6.1 Experiment Test

The participant samples consisted of students who were registered for CPCS 351 at the Department of Computer Science in the FCIT of KAU, during Fall 2017 and Fall 2018. The students in the Fall 2018 group sat the final exam with the blueprint, whereas the students in the Fall 2017 group sat the final exam without the blueprint. Both exams were designed by the course coordinator, who had taught the same course for more than three years. The independent variables in the study comprised the number of items in each type of question that aligned to the selected ILO (Fig. 3). The dependent variables comprised the performance of the students.

6.2 Test Environment

All the participants sat the final exam in a closed room on the second floor of the FCIT building in KAU. Each participant who sat the exam in Fall 2017 received the same exam, and each participant who sat the exam in Fall 2018 received the same exam. The final exam for Fall 2017 was prepared without using the blueprint, and the final exam for Fall 2018 was prepared using the blueprint and included the questions formulated according to the ILOs.

6.3 Hypothesis

A set of hypotheses, one related to the Fall 2017 and one related to the Fall 2018, was formulated and investigated empirically. The hypotheses are listed below.

HYP0: There is no significant difference between the performance of the students of Fall 2017, who did not use a blueprint exam, and the performance of the students of Fall 2018, who did use a blueprint exam.

HYP1: There is a significant difference between the performance of the students of Fall 2017, who did not use a blueprint exam, and the performance of the students of Fall 2018, who did use a blueprint exam

7. Results

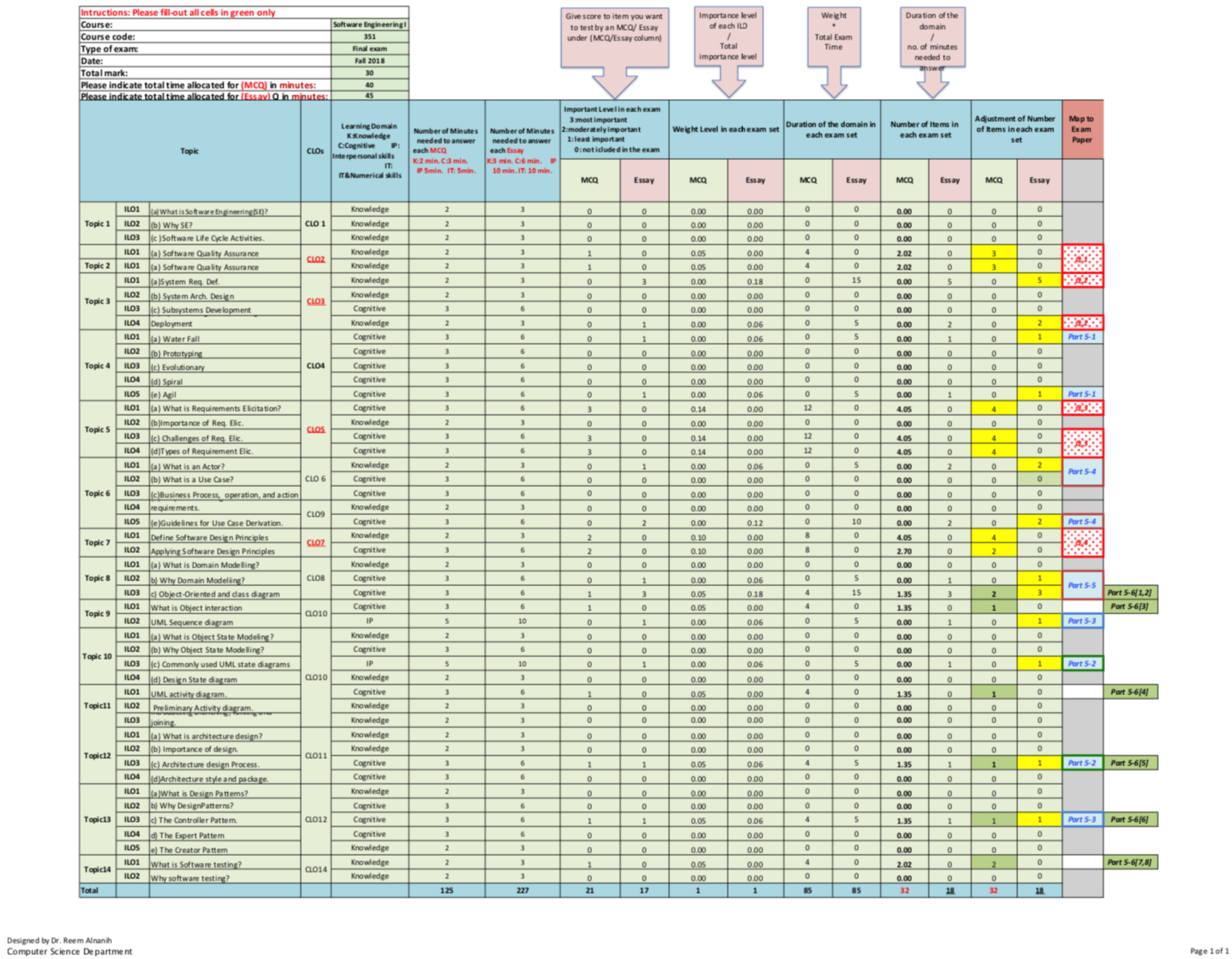
The performances of 39 students in Fall 2017 and 64 students in Fall 2018 were measured, and the performance measurement comprises the objective factor of the study.

The raw data for the study were formulated in Microsoft Excel for each group. The final exam was calculated from 30 marks for each group. We calculated the dependent variables (the result of each item) for each group. Then, we used the mean of all the students in each group to compare the performance of the students in Fall 2017 and Fall 2018. The data are presented in two separate tables. For each group, since mean and median of the data are the same (Table 2), the data are normal, and we have two different samples—39 students and 64 students. Consequently, we used an analysis of variance (ANOVA) test for the data analysis. Since we are 95% confident that the mean of each group lies within the confidence interval, we relied on the mean.

Table 2: The Mean and Median for Group 1 and Group 2

	Group 1 Fall 2017	**Group 2 Fall 2018**
Mean	22.05	23.3
Median	22	23

An ANOVA is a statistical technique used to check if the means of two or more groups are significantly different. An ANOVA determines the impact of one or more factors by comparing the means of different samples [26]. Based on the mean (Table 2), the mean grades of Group 1 (Fall 2017) were not the same as Group 2 (Fall 2018).



Intructions: Please fill-out all cells in green only	
Course:	Software Engineering I
Course code:	351
Type of exam:	Final exam
Date:	Fall 2018
Total mark:	30
Please indicate total time allocated for (MCQ) in minutes:	40
Please indicate total time allocated for (Essay) Q in minutes:	45

Topic			CLOs	Learning Domain K:Knowledge C:Cognitive IP: Interpersonal skills IT: IT&Numerical skills	Number of Minutes needed to answer each MCQ K:2 min. C:3 min. IP 5min. IT: 5min.	Number of Minutes needed to answer each Essay K:3 min. C:6 min. IP 10 min. IT: 10 min.	Important Level in each exam 3:most important 2:moderately important 1:least important 0:not icluded in the exam		Weight Level in each exam set		Duration of the domain in each exam set		Number of Items in each exam set		Adjustment of Number of Items in each exam set		Map to Exam Paper	
							MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay	MCQ	Essay		
Topic 1	ILO1	(a)What is Software Engineering(SE)?	CLO 1	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO2	(b) Why SE?		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c)Software Life Cycle Activities.		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO1	(a) Software Quality Assurance	CLO2	Knowledge	2	3	1	0	0.05	0.00	4	0	2.02	0	3	0	Q.1	
Topic 2	ILO1	(a) Software Quality Assurance		Knowledge	2	3	1	0	0.05	0.00	4	0	2.02	0	3	0		
Topic 3	ILO1	(a)System Req. Def.	CLO3	Knowledge	2	3	0	3	0.00	0.18	0	15	0.00	5	0	5	Q.2	
	ILO2	(b) System Arch. Design		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c) Subsystems Development		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO4	Deployment		Knowledge	2	3	0	1	0.00	0.06	0	5	0.00	2	0	2	Q.2	
Topic 4	ILO1	(a) Water Fall	CLO4	Cognitive	3	6	0	1	0.00	0.06	0	5	0.00	1	0	1	Part 5-1	
	ILO2	(b) Prototyping		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c) Evolutionary		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO4	(d) Spiral		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO5	(e) Agil		Cognitive	3	6	0	1	0.00	0.06	0	5	0.00	1	0	1	Part 5-1	
Topic 5	ILO1	(a) What is Requirements Elicitation?	CLO5	Cognitive	3	6	3	0	0.14	0.00	12	0	4.05	0	4	0	Q.3	
	ILO2	(b)Importance of Req. Elic.		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c) Challenges of Req. Elic.		Cognitive	3	6	3	0	0.14	0.00	12	0	4.05	0	4	0	Q.3	
	ILO4	(d)Types of Requirement Elic.		Cognitive	3	6	3	0	0.14	0.00	12	0	4.05	0	4	0		
Topic 6	ILO1	(a) What is an Actor?	CLO 6	Knowledge	2	3	0	1	0.00	0.06	0	5	0.00	2	0	2	Part 5-4	
	ILO2	(b) What is a Use Case?		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c)Business Process, operation, and action		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO4	requirements.	CLO9	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO5	(e)Guidelines for Use Case Derivation.		Cognitive	3	6	0	2	0.00	0.12	0	10	0.00	2	0	2	Part 5-4	
Topic 7	ILO1	Define Software Design Principles	CLO7	Knowledge	2	3	2	0	0.10	0.00	8	0	4.05	0	4	0	Q.4	
	ILO2	Applying Software Design Principles		Cognitive	3	6	2	0	0.10	0.00	8	0	2.70	0	2	0		
Topic 8	ILO1	(a) What is Domain Modelling?	CLO8	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO2	b) Why Domain Modeling?		Cognitive	3	6	0	1	0.00	0.06	0	5	0.00	1	0	1	Part 5-5	
	ILO3	c) Object-Oriented and class diagram		Cognitive	3	6	1	3	0.05	0.18	4	15	1.35	3	2	3		Part 5-6[1,2]
Topic 9	ILO1	What is Object interaction	CLO10	Cognitive	3	6	1	0	0.05	0.00	4	0	1.35	0	1	0		Part 5-6[3]
	ILO2	UML Sequence diagram		IP	5	10	0	1	0.00	0.06	0	5	0.00	1	0	1	Part 5-3	
Topic 10	ILO1	(a) What is Object State Modeling?		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO2	(b) Why Object State Modelling?		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c) Commonly used UML state diagrams		IP	5	10	0	1	0.00	0.06	0	5	0.00	1	0	1	Part 5-2	
	ILO4	(d) Design State diagram	CLO10	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
Topic11	ILO1	UML activity diagram.		Cognitive	3	6	1	0	0.05	0.00	4	0	1.35	0	1	0		Part 5-6[4]
	ILO2	Preliminary Activity diagram.		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	joining.		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
Topic12	ILO1	(a) What is architecture design?	CLO11	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO2	(b) Importance of design.		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	(c) Architecture design Process.		Cognitive	3	6	1	1	0.05	0.06	4	5	1.35	1	1	1	Part 5-2	Part 5-6[5]
	ILO4	(d)Architecture style and package.		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
Topic13	ILO1	(a)What is Design Patterns?	CLO12	Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO2	b) Why DesignPatterns?		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO3	c) The Controller Pattern.		Cognitive	3	6	1	1	0.05	0.06	4	5	1.35	1	1	1	Part 5-3	Part 5-6[6]
	ILO4	d) The Expert Pattern		Cognitive	3	6	0	0	0.00	0.00	0	0	0.00	0	0	0		
	ILO5	e) The Creator Pattern		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
Topic14	ILO1	What is Software testing?	CLO14	Knowledge	2	3	1	0	0.05	0.00	4	0	2.02	0	2	0		Part 5-6[7,8]
	ILO2	Why software testing?		Knowledge	2	3	0	0	0.00	0.00	0	0	0.00	0	0	0		
Total					125	227	21	17	1	1	85	85	32	18	32	18		

Fig. 3 The blueprint for CPCS 351 for Fall 2018.

Table 3: ANOVA Test Results of Group 1 and Group 2

ANOVA Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Group 1	39	860	22.0512821	7.59927463
Group 2	64	1504.5	23.5078125	10.1685888

ANOVA

Source of variation	SS	Df	MS	F-value	P-value	F-critical value
Between groups	51.4098684	1	51.4098684	5.58686557	0.02000927	3.93518869
Within groups	929.39353	101	9.20191614			
Total	980.803398	102				

For statistical validation, Table 3 presents the ANOVA Test for Group 1 and Group 2. From the above table, we can see that the F-value is greater than the F-critical value for the alpha level selected (0.05). Therefore, there is evidence to reject the null hypothesis, and there is a significant difference between Group 1 and Group 2. Also, since the P-value was less than the alpha level (0.05), the null hypothesis was again rejected, and the alternative hypothesis was accepted.

In order to investigate which group performed better in the final exam, a t-test was used. Two samples assuming unequal variances were applied to the data of Group 1 and Group 2.

Table 4: t-test: Two samples assuming unequal variances

	Group 1	Group 2
Mean	22.05128205	23.3382353
Variance	7.599274629	10.5462028
Observations	39	68
Hypothesized mean difference	0	
Df	90	
t-Stat	-2.175520781	
P(T<=t) one-tail	0.016105818	
t critical one-tail	1.661961084	
P(T<=t) two-tail	0.032211637	
t critical two-tail	1.986674541	

Table 4 confirms that the P-value (0.03) is less than the alpha (0.05). Since the t-test values obtained for the performance lie in the critical region, the null hypothesis for the performance factor is rejected. Therefore, HYP1 is accepted meaning that there is a significant difference between two groups, and based on the mean, the performance of Group 2 (Fall 2018) is better than that of Group 1 (Fall 2017).

8. Discussion

During Fall 2017, CPCS 351 was assessed through direct assessment methods (i.e., two exams, two assignments, three graded labs, one final exam, and one term project) and an indirect assessment (i.e., a survey) that was distributed by the course coordinator to the students. Four high course learning outcomes (HCLOs) were assessed in the comprehensive final exam. All the students in Group 1 sat the same final exam. The average grade for the final exam was 74%. The attainment rates for the four HCLOs (CLO #2, CLO #3, CLO #5, and CLO #7) in the final exam were as follows: 93.5%, 69.76%, 98.3%, and 82.45%, respectively. The average rate for all HCLOs was 86%.

During Fall 2018, CPCS 351 was assessed through direct assessment methods (i.e., two exams, two assignments, three graded labs, one final exam, and one term project) and an indirect assessment (i.e., a survey) that was distributed by the course coordinator to the students. Four HCLOs were assessed in the comprehensive final exam. All students in Group 2 sat the same final exam. The average grade for the final exam was 78%. The attainment rates for the four HCLOs (CLO #2, CLO #3, CLO #5, and CLO #7) in the final exam were 82%, 92.5%, 94.5%, and 89.5%, respectively. The average rate for all HCLOs was 90%.

Overall, the above results are satisfactory for Fall 2018, especially in relation to CLO #3, which increased from 69.76% in Fall 2017 to 92.5% in Fall 2018.

For the indirect assessment, the coordinator designed and distributed the survey to the students. The results showed very positive feedback for Fall 2018.

A study can provide ambiguous results in many ways. The classification of the potential problems with empirical studies is based on the threats these problems pose to the study's validity. The following factors describe the four categories of validity threats [27]-[28]:

1) Conclusion Validity: There are two main objectives of our experiment—namely, to confirm a theory of proposing and adapting blueprint in a final exam (Section 5), by applying an ANOVA test to our data (Section 6), and to explore the relationships between datasets using a t-test to confirm if a relationship exists between student performance and the application of a blueprint to an exam. Through implementation, we generated measures of association which indicate the closeness of two variables.

2) Construct Validity: A theoretical validation of all attributes has been provided for the proposed blueprint, meaning that for a valid measure, it should reflect the real meaning of the concept under consideration (Section 5). The independent variables in the study comprised the number of items in each type of question that aligned with

the selected ILO (Fig. 3). The dependent variable comprised the performance of the students.

3) Internal Validity: A study is internally valid if the treatment leads to the effect depicted in the dependent variables. The independent variables (i.e., the number of items in each type of question) affect the dependent variable (i.e., the performance of the students) if any kind of manipulation of the independent variables creates an impact on the outcome. The dependent variable supports the outcome since it is affected by updating the value of one or more independent variables.

4) External Validity: The results of the controlled experiment are applicable and extendable to anyone familiar with blueprint evaluation method. As well as this, it can be applied in and considered by all departments of the FCIT in KAU.

9. Conclusion

The purpose of this paper was to design an adaptable test-based blueprint for a software engineering course in the Department of Computer Science at the FCIT in KAU. The research defined the structure and guidelines needed to design the blueprint for an exam. The importance of using the blueprint in different assessment methods was also confirmed, as the blueprint not only clarifies the assessment but also helps the instructor to determine, both subjectively and objectively, the adequate coverage of content and ensure emphasis is placed where intended.

The main goal of assessment-based exams is to accurately measure student performance in terms of their required knowledge and skills, which are generally tabulated within learning objectives.

A well-constructed blueprint is a versatile tool that can enhance all aspects of the course design and evaluation. An initial investment of time and effort in creating a blueprint that amalgamates the learning objectives, learning experiences, and evaluation pays off in the long run. After creating a blueprint, the systematic monitoring of content must be maintained to achieve content validity. Blueprint transparency drives instructors' teaching and students' learning [15].

Effective blueprinting depends on the level of organization of the course. The proposed blueprint presented in this paper is based on an undergraduate computer science course at KAU. The experimental test conducted on the two groups of students confirms that there is a difference when a blueprint is applied versus when it is not. The results confirm that applying a blueprint will significantly improve student performance.

Since blueprinting in assessments provides a fair, valid, and reliable evaluation of students, there is an urgent need for the instructors at the Department of Computer Science to become familiar with blueprinting in assessments in order to achieve uniform, reliable, and valid assessments.

The blueprint was designed with consideration of the following specific, measurable, achievable, relevant, and time-based (SMART) goals:

- Specific ILOs that have only one clear meaning.
- Measurable ILOs that can be expressed and converted into values.
- An achievable model that has to be implemented within the available domains, activities, and methods.
- A model that is relevant to the subject, to students, to the program's objectives, and to the institution's mission.
- A timely model that has to be implemented within a specific timeframe.

The hope is that this paper will improve the quality of written exams. The future directions of this work include broad testing of the proposed blueprint in information technology departments and information systems at FCIT.

References

[1] F. E. Go, "Blueprint," in Encyclopedia Britannica 3 (Expo'70 edition). Chicago, IL, USA: William Benton, p. 816, 1970.

[2] M. N. Woods, *From Craft to Profession: The Practice of Architecture in Nineteenth-Century America*. UCP, 1999.

[3] R. L. O. Pinto and F. Maloberti, "X RAY and blueprint: Tools for MOSFET analog circuit design addressing short-channel effects," 2004 IEEE International Symposium on Circuits and Systems. Vancouver, Canada: IEEE, pp. V277–V280, 2004.

[4] M. Najjar, "A blueprint practical implementation of PKI using open PGP at University of Tabuk," paper presented at the Science and Information Conference, London, Oct. 2013.

[5] H. Hamdy, "Blueprinting for the assessment of health care professionals," *Clin Teach*, vol. 3, no. 3, pp. 175–179, Sep. 2006.

[6] D. Newble et. al., "Guidelines for assessing clinical competence," Teach Learn Med, vol. 6, pp. 213–220, 1994.

[7] A. I. Gonzalez-Tablas Ferreres, K. Wouters, B. Ramos Alvarez, and A. Ribagorda Garnacho, "EVAWEB: A web-based assessment system to learn X.509/PKIX-based digital signatures," IEEE Trans Education, vol. 50, no. 2, pp. 112–117, May 2007.

[8] E. Guzman and R. Conejo, "Self-assessment in a feasible, adaptive web-based testing system," IEEE Trans Education, vol. 48, no. 2, pp. 688–695, Nov. 2005.

[9] National Commission for Academic Accreditation and Assessment. Handbook for Quality Assurance and Accreditation in Saudi Arabia: Part 2.Version 3, October 2015.
http://www.kfupm.edu.sa/deanships/dad/Documents/AAC/NCAAA%20Documents/H2.%20Handbook%20Part%202.pdf

[10] National Qualifications Framework for Higher Education in the Kingdom of Saudi Arabia. 2009.

https://www.mu.edu.sa/sites/default/files/National%20Qualifications%20Framework%20for%20HE%20in%20KSA.pdf

[11] K. McLaughlin, S. Coderre, W. Woloschuk, and H. Mandin, "Does blueprint publication affect students' perception of validity of the evaluation process?" Adv Health Sci Educ, vol. 10, pp. 15–22, 2005.

[12] P. Bridge, J. Musial, R. Frank, T. Roe, and S. Sawilowsky, "Measurement practices: Methods for developing content-valid student examinations," Med Teach, vol. 25, no. 4, pp. 414–421, 2003.

[13] D. Sales, A. Sturrock, K. Boursicot, and J. Dacre, "Blueprinting for clinical performance deficiencies – Lessons and principles from the General Medical Council's fitness to practice procedures," Med Teach, vol. 32, pp. e111–e114, 2010.

[14] R. G. Ahmad and O. A. E. Hamed, "Impact of adopting a newly developed blueprinting method and relating it to item analysis on students' performance," Med Teach, vol. 36, supplement 1, pp. S55–S61, April 2014.

[15] S. Y. Patil, M. Gosavi, H. B. Bannur, and A. Ratnakar, "Blueprinting in assessment: A tool to increase the validity of undergraduate written examinations in pathology," Int J Appl Bas Med Res, vol. 5, supplement 1, pp. S76–S79, 2015.

[16] L. R. Becker and M. Vassar, "An assessment blueprint for the Advanced Medical Life Support two-day prehospital emergency medical services training program in the United States," J Educ Eval Health Prof, pp. 1–8, 2015.

[17] H. Abdellatif and A. M. Al-Shahrani, "Effect of blueprinting methods on test difficulty, discrimination, and reliability indices: Cross-sectional study in an integrated learning program," Adv Med Educ Pract, vol. 10, pp. 23–30, 2019.

[18] T. Patel, M. K. Saurabh, and P. Patel, "Perceptions of the use of blueprinting in a formative theory assessment in pharmacology education," Sultan Qaboos Univ Med J, vol. 16, no. 4, e475, 2016.

[19] S. S. Kadam, S. K. Kolhe, and V. Kulkarni, "Knowledge, attitude and practices about use of blueprinting in assessment among medical teachers," Indian J Appl Res, vol. 9, no. 8, 2019.

[20] K. M. A. Salih et al., "Students perception towards blueprint in the College of Medicine, University of Bisha, KSA," MRJER, vol. 6, no. 5, pp. 79–82, May 2018.

[21] S. Coderre, W. Woloschuk, and K. McLaughlin, "Twelve tips for curriculum blueprinting," *Med Teach*, vol. 31, pp. 322–324, 2009. DOI: 10.1080/01421590802225770.

[22] T. M. El-Gohary, "Dr. El-Gohary blueprint matrix: A feasible template purported to be a quick and simple guide for academics to design robust multiple-choice questions exams," *IJTRR*, vol. 6, pp. 38–44, 2017.

[23] I. S. Alfallay, "Test specifications and blueprints: Reality and expectations," *Int J Instr*, vol. 11, no. 1, pp. 195–210, 2018. DOI: 10.12973/iji.2018.11114a

[24] E. M. Gaffas, R. P. Sequeira, R. A. Al Namla, and K. S. Al-Harbi, "Test blueprints for psychiatry residency in-training written examinations in Riyadh, Saudi Arabia," *Adv Med Educ Prac*, vol. 3, pp. 31–46, 2013.

[25] M. J. Rudolph, K. K. Daugherty, M. E. Ray, V. P. Shuford, L. Lebovitz, and M. V. DiVall, "Best practices related to examination item construction and post-hoc review," Am J Pharm Educ, vol. 83, no. 7, 7204, 2019.

[26] J. Rojewski, I. H. Lee, and S. Gemici, "Use of t-test and ANOVA in career-technical education research," *CTER*, vol. 37, no. 3, pp. 263–275, 2012.

[27] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering: An Introduction*, Norwell, MA, USA: Kluwer, 2000.

[28] N. Fenton and J. Bieman, *Software Metrics: A Rigorous and Practical Approach* (Third Edition), Boca Raton, FL, USA: CRC Press Inc., 2014.

Reem Alnanih is currently associated with the Department of Computer Science in the Faculty of Computing and Information Technology (FCIT) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia. She received the Ph.D. degree in Computer Science from Concordia University, Montreal, Canada, in 2015. She obtained the M.S. degree in Computer Science from KAU, Jeddah, Saudi Arabia, in 2008. She is head of the Innovation and Entrepreneurship Unit of the FCIT at KAU. Her research interests include human–computer interaction, user experience, software engineering, quality measurement and associated evaluation techniques, and the Internet of Things.