

Article

Reality vs. Expectations of Assessment in STEM Education: An Exploratory Case Study of STEM Schools in Egypt

Mohamed Ali El Nagdi ^{1,*}  and Gillian H. Roehrig ²¹ Department of Educational Studies, The American University in Cairo, New Cairo 11835, Egypt² Department of Curriculum and Instruction, University of Minnesota, St Paul, MN 55108, USA

* Correspondence: mohamedelnagdi@aucegypt.edu

Abstract: In this exploratory case study, the assessment methods planned and used in Egyptian STEM schools were explored. The purpose of the study was to explore the relationship between the ideals provided in STEM education both from research and policy documents and the actual assessment strategies used both at the classroom and state level in order to understand the alignment between the proposed lofty goals of STEM and the modes of assessment actually used. Teachers in Egyptian STEM schools were surveyed and interviewed to explore this relationship. Samples of their assessments were also examined. Teachers were found to have been using two mutually exclusive models of assessment; a set of assessments at the disciplinary level and another set at multidisciplinary level including, but not restricted to, project and problem learning, inquiry, and reflective journaling. The study revealed partial alignment between expectations and reality of assessment in Egyptian STEM schools.

Keywords: STEM education; assessment; constructive alignment; twenty-first century skills



Citation: El Nagdi, M.A.; Roehrig, G.H. Reality vs. Expectations of Assessment in STEM Education: An Exploratory Case Study of STEM Schools in Egypt. *Educ. Sci.* **2022**, *12*, 762. <https://doi.org/10.3390/educsci12110762>

Academic Editors: Lihua Xu, Su-Chi Fang and Wany Widjaja

Received: 27 September 2022

Accepted: 25 October 2022

Published: 28 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As science, technology, engineering, and mathematics (STEM) education continues to gain momentum as an educational reform, both at policy and curriculum levels [1], it is high time to explore the assessment systems used in different STEM settings to understand if and how STEM approaches are delivering on their promise. Advocates argue that STEM provides a vehicle for preparing the 21st century workforce, in terms of both necessary skills and knowledge [2]. Effective STEM learning requires that both assessment and learning processes are integrated by both teachers and curriculum developers in an organic and complementary relationship [3–6]. However, difficulties exist because of tensions between classroom assessments whose primary aim is students' learning and intellectual development and standardized tests that target ranking students [7]. While the relationship between classroom assessment and students' learning is clear, there is less clarity in the case of large-scale standardized assessment because "it is mediated through policy, curriculum, and assessment design" [7], p. 340.

While current policies call for students to learn through hands-on, project based, inquiry-based teaching and learning approaches in STEM classrooms [8,9], they are still being assessed using traditional assessment systems mostly through standardized tests unaligned with the goals of integrated STEM [10–13]. This seemingly poor alignment between policy goals for STEM and current assessment practices [14] warrants further research. This study explores methods of assessment for STEM in Egypt as a specific case and is guided by the following research questions:

RQ1: What are the different assessment methods used in STEM schools in Egypt?

RQ2: In what way, if any, do the teachers describe assessment systems used in STEM schools as different from the mainstream assessment systems and standardized tests?

RQ3: How are teachers' STEM assessment practices aligned with what STEM advocates? What are the challenges that hinder this alignment?

2. Literature Review

2.1. STEM Education

STEM education addresses the increasing needs for a qualified STEM workforce by providing a solution to students' lack of interest in STEM careers and targeting the development of the skills necessary for STEM careers, especially for minority and female students [15–17]. However, while students are expected to deepen their knowledge of STEM [18], STEM education should not be limited to promoting further study in STEM; the development of STEM literacy is equally critical [19–21] to becoming a productive global citizen equipped with the 21st century skills of creativity, communication, collaboration, and critical thinking [22]. Such lofty aspirations require dedication on the part of all stakeholders, notably teachers, who are expected to develop robust curriculum designs that align assessment with the intended learning outcomes [11,23].

However, an ongoing challenge faced by STEM educators is the lack of consensus on definitions of STEM [8,24]. This is critical in understanding assessment in STEM, as it matters if STEM is a unified body of knowledge on which assessment can be anchored or a pedagogical approach to actualize the aforementioned goals of STEM education [25,26]. Nonetheless, there are efforts to put forward a framework for STEM that may help researchers and practitioners to reach common ground on the critical features of integrated STEM. In a recent review, Roehrig et al. [22] put forth a detailed conceptual framework for K-12-integrated STEM education generated from the extant integrated STEM literature. Built on consensus areas within the literature, they proposed seven central characteristics of integrated STEM: (a) centrality of engineering design, (b) driven by authentic problems, (c) context integration, (d) content integration, (e) STEM practices, (f) twenty-first century skills, and (g) informing students about STEM careers. However, absent from the discussion about definitions and conceptualizations of STEM is an explicit connection to assessment.

While teachers at STEM schools in Egypt see the introduction of STEM education as a great breakthrough in the Egyptian education system, they still have a blurry understanding of what it really means, especially in terms of properly assessing their students [27]. With the different ways STEM is conceptualized by individual teachers, implementation varies depending on what STEM means and how this is interpreted in practice. However, the demand for STEM education and research reveals a need to move from the traditional disciplinary teaching and learning to a more integrated system [28] where students make sense of what they learn in school. However, teachers and students in STEM schools still grapple with the challenge of how to get out of the vicious circle of seeing assessment as simply an evaluation of rote learning [27,29].

2.2. Assessment for or of Learning

Traditionally, students have been assessed or rather evaluated at the end of their learning journeys, termed as summative assessment, for purposes that may not be linked to learning but to ranking and making sure that the intended learning outcomes have been mastered [30]. This approach to assessment has been seen as inefficient to both guide instruction and support learning. As early as 1967, the term formative assessment was introduced by Michael Scriven to refer to assessment efforts directed towards helping students see their levels of performance more than once before sitting for a final round of exams. This resulted in doing assessment more frequently and analyzing the data gathered over several exams for better instruction and support for students [30]. However, the type of assessment methods used in this formative assessment were more or less standardized tests, mostly multiple-choice questions whose results are analyzed and used to improve both learning and instruction but still fall short of catering for the diversity of students' learning profiles, interests, and readiness to learn [31]. Therefore, more performance-based authentic assessment methods have been introduced to cater for assessing the learner as a whole as well as the learning and teaching process where authenticity is seen as contextualizing and problematizing the learning process to link what happens in the classroom with real life [32–34].

2.3. Assessment in STEM

While STEM education aims to motivate students' learning of STEM content and promote 21st century skills, developing assessment of such interdisciplinary learning remains a challenge [35–37]. International standardized assessments such as TIMMS and PISA continue to measure students' disciplinary knowledge rather than cross-disciplinary knowledge and skills. With this in mind, exploring and scrutinizing assessment frameworks in STEM education should be a priority at both research and policy levels.

2.4. STEM Assessment Frameworks

Despite the goals of current STEM reforms, research continues to describe student achievement in STEM by using scores in standardized assessments in mathematics and science [38–40]. This reliance on available standardized test data in science and mathematics has led to the development of assessment frameworks that purport to relate to integrated STEM but do not address other learning goals relevant to STEM education such as skills and attitudes [41]. For example, Bicer, Capraro, and Capraro [42] developed an integrated STEM assessment model based on a second-order confirmatory analysis of science and mathematics scores of 231,966 students, yielding a model with an adequate fit to the data. However, what STEM promises is not just improving test scores in mathematics and science; it also includes, among other things, developing social responsibility, autonomy, collaboration, design thinking (especially in engineering design), and problem solving [20,43].

To this end, Arikan and colleagues [44] took a different approach, using the literature to design an assessment framework for STEM that is based on “non-routine” problem solving that takes into account the multidimensional and integrated nature of STEM competencies of algorithm thinking, concept and principles, pattern recognition, argumentation, science literacy, and engineering and technology problems. These competencies were used as subdomains for each of the four STEM domains (science, technology, engineering, and mathematics) and they developed a test based on this framework that was calibrated using item response theory. The authors relied on science, technology, engineering, and mathematics-related problems that required mathematical calculations as a medium to develop an assessment tool for this framework. While their empirical findings provided evidence of adequate validity for the structure of their developed framework in which the subdomains are interrelated to one another, the study has yet to generate generalizability of the results, especially in more diverse populations.

Looking at STEM education as a discipline in itself, in addition to being contentious, contradicts the very goals for introducing the approach as a framework for integration and holistic learning that develops 21st century skills [22]. Therefore, restricting assessment tools in STEM education to merely yield appropriate measures of concepts from the STEM disciplines is controversial. The challenge often lies in how teachers assess students' scientific and STEM literacy skills, aptitude, as well as other intangible skills in parallel to gains in content knowledge in paper-and-pencil tests [35]. This tension is still one of the major challenges to the vision of STEM education as an integrated approach to improve student learning of the STEM disciplines using alternative assessment techniques interwoven into the instructional pedagogies has been one of the features of successful STEM teaching and learning. This argument implies that assessment in the context of STEM education should encompass both disciplinary competency [44,45] and competencies that are attributed to more than one STEM discipline such as project-based learning skills where students design a solution to problems following an engineering design process [43].

2.5. Assessment Strategies in STEM

While the structure and epistemological grounding for STEM assessments have yet to be established, the literature presents a number of studies that describe assessment strategies used by teachers in STEM classrooms. For example, Gao and colleagues [36] conducted a systematic review of literature on the assessment used by teachers of student learning in

integrated STEM education. They used a two-dimensional framework which comprised (1) nature of the multiple disciplines to be assessed and (2) the learning objectives in relation to the multiple disciplines. They argued that most assessments are monodisciplinary wherein the emphasis is on the subject the teacher teaches, with emphasis on knowledge and affective (attitude) domains. At the far end of the spectrum, transdisciplinary STEM assessment emphasizes the measurement of the affective domain. They concluded that although many programs aimed to improve students' interdisciplinary understanding or skills, their assessments did not align with their aims. Septiani and Rustaman [46] found out that the use of performance assessment in a STEM setting is more effective in detecting students' science process skills than individual observation when compared through science process skills (SPS) test.

3. Methods

3.1. Purpose of the Research

The purpose of this study was to explore the assessment methods used in STEM education settings in Egypt. Participants were asked to describe their actual assessment practices which could then be examined for alignment with the expectations for STEM education in Egypt.

3.2. Research Design

To this end, an exploratory case study was utilized [47]. The use of the exploratory case study is appropriate for the research at hand since it is controlled by the boundaries of the Egyptian STEM schools, and it is exploratory in nature as exploring assessment in STEM settings in Egypt is unique. A sequential mixed-method data collection approach was used to provide both quantitative and qualitative data [48]. An online survey using Google forms in which participants described their assessment techniques, shared their schools' vision and mission and provided samples of their work, was followed by semi-structured qualitative interviews with purposefully selected participants. This mixed-method approach provides depth and detail to a study and potentially uncovers new insights into participant experiences [49].

3.3. Context (Setting)

In 2011, Egypt's STEM reforms started with the opening of the first STEM school; currently 19 STEM schools are in operation. The Egyptian experience is influenced by the US STEM experience due to different factors including, but not limited to, the expertise, initial funding, assistance in curriculum development, and professional development provided through an educational consortium funded by the USAID [27,50,51]. The educational consortium was a group of US education companies and experts and supported by Egyptian experts supported by Egyptian staff on the ground. Running schools and recruiting teachers was the MoE&TE's responsibility. Five years later when the first grant was concluded, all responsibilities including professional development had been turned over to the MoE&TE [50]. Teaching in the STEM schools is based on project-based learning and supported by extra-curricular research opportunities in partnership with universities and research centers. STEM Schools in Egypt are public schools managed by the STEM Central Unit in the MoE&TE, and hence all schools follow the same curriculum and assessment system. Assessment occurs at the subject level to monitor students' progress in both learning the subject and project work. At the disciplinary level and according to the ministerial decrees 382/2012 [52] and 238/2013 [53], the final exams for the disciplinary content areas are designed in the central MoE&TE STEM unit, and the grading of the final exams is performed centrally while the capstone final exam (exhibition of the final projects) is held on school premises and evaluated by external judges (See Table 1). The capstone project is an interdisciplinary project where students work on solving one of the grand challenges of Egypt, mostly in groups. In contrast, the final exams are standardized university readiness tests (URTs) and concept tests (CTs) held at the end of the school year.

Goals

The Ministry of Education and Technical Education Egypt has started the STEM education experience with great hopes to reform the struggling education system as a whole and open horizon for Egyptian youth to adopt 21st century skills and pursue STEM field studies as they decide on higher education routes [54]. Several Ministerial decrees elucidated this vision including Ministerial decree number 382/2012 that stated the goals for establishing STEM schools in Egypt are as follows:

- 1: Students must demonstrate a deep understanding of the scientific mathematical and social dimensions of Egypt's grandest challenges as a country.
- 2: Students must demonstrate understanding of the content and ways of knowing that display scientific, mathematical and technological literacy and subject matter proficiency.
- 3: Students must exhibit self-motivation, self-direction and a hunger for continued learning.
- 4: Students must exhibit the ability to think independently, creatively and analytically.
- 5: Students must exhibit the ability to question, collaborate and communicate at a high level.
- 6: Students must demonstrate the capacity to become socially responsible leaders
- 7: Students must be to apply their understanding to advance creativity, innovation and invention with a real world vision with a consciousness and eye toward a more contemporary Egypt.
- 8: Goals 1–7 must be implemented and viewed through the lens of a Digital Platform. Students must become fluid with technology to ensure that they maximize digital methods of data storage and communication [54], p. 45.

Table 1. Assessment system in Egyptian STEM schools as set up by the Ministry of Education and Technical Education (MoE&TE) in Egypt.

Assessment Structure
Grades 10 and 11 according to ministerial decree number 382/2012
Students will be awarded a total score, which will be calculated based on four different indicators, as follows:
(A) Final exam with special requirements (midterm exams) 30%;
(B) Capstone projects 60%;
(C) Attendance and participation 10%.
Grade 12 according to ministerial decree number 238/2013
The secondary STEM certificate is limited to 3rd year examinations.
The committee in charge of developing the final year examinations are listed as follows:
(A) One subject area advisor;
(B) One expert in STEM Education;
(C) Two professors from universities and think tanks to be nominated by the Scientific Research Academy, among other parties, at the discretion of the Board of Directors.
Students will be awarded a total score, which will be calculated based on four different indicators, as follows:
(A) University Readiness Test (40%) (very similar to the ACT exam);
(B) Measures/Inventories of concepts achieved by students in science and mathematics (20%);
(C) Student performance in Capstone projects (20%);
(D) Student attendance and participation, which shall constitute ten percent (10%) of the total score. This shall be assessed and evaluated by various subject matter teachers under the supervision of school directors;
(E) The remaining 10% is awarded to presentation and lab work in science and math 5% each, but for humanities the whole 10% is awarded to the research and presentation work done throughout the whole academic year.

Based on MoE&TE decrees 382/2012 and 238/2013.

3.4. Participants

In this study, 22 teachers at STEM Schools in Egypt; mainly from MoE&TE STEM schools (19 teachers), with a few from self-proclaimed private STEM schools (3 teachers)

participated in this study after an invitation was sent to them via email and WhatsApp message. Demographic information is provided in Table 2.

As a follow-up to the survey, six MoE&TE STEM School teachers were purposefully selected to be interviewed as these schools provide a robust context and in-depth insights into the assessment system utilized in Egyptian STEM schools (see Table 3). They were purposefully selected due to different factors: their teaching experience in the Egyptian STEM schools and their non-STEM teaching experience as well; their representation of a variety of subjects and schools; and finally, their roles in these schools. Internal Review Board approval as obtained, and teachers signed consent forms to participate in this study prior to conducting any semi structured interviews.

Table 2. Participant Demographics.

Total Number of Participants = 22	
Number of teachers from MoE&TE STEM schools	19
Number of teachers from other aspiring STEM schools	3
Subjects	
Science	13 (biology–chemistry–physics)
Technology	1 computer science
Engineering	2 leading STEAM projects
Arts	1 music
Mathematics	2
Language arts	3 (2 English and 1 homeroom KG teacher)
Range of teaching experience in STEM	2 to 7 years
Overall teaching Experience	5 to 28 years

Table 3. Interviewed Participants.

Name *	School *	Subject	Experience in Teaching	Experience in STEM Teaching	Comments
Ebtisam	STEM School 1	Biology	More than 20 years	10 years	Capstone Coordinator
Hadeer	STEM School 2	Chemistry	15 years	5 years in 2 STEM schools	
Omar	STEM School 3	Computer Science	18 years	5 years	
Dalila	STEM school 4	English	20 years	3 years	
Kareem	STEM School 5	Biology	10 years	5 years	Capstone coordinator
Mona	STEM School 6	Science	10 years	One year	

* The names of teachers and schools are pseudonyms.

4. Data Collection

4.1. Survey

An online survey was created with a set of questions to explore teachers' assessment use in their classrooms and how these choices reflect the goals of STEM education (See Appendix A). The survey asked teachers to describe the types of assessment they use in their classes, as well as their schools' vision and mission, and how these assessments are different from mainstream education systems. They were also asked to share examples and artifacts of assessments from their classes.

4.2. Semi Structured Interviews

The six purposefully selected MoE&TE STEM School teachers were interviewed in order to provide a robust context and in-depth insights into the assessment system utilized in their STEM schools. In the interviews (see Appendix B), we further explored these teachers' perspectives and understanding of the assessment systems they use. The interview questions were designed to both align with the research questions and to probe both

the teachers' understanding of assessment in general, assessment in STEM, the tools and techniques used, challenges, and how it is different from assessment in mainstream education. The interviews were conducted through Zoom mainly in Arabic, recorded and analyzed later by the research team.

5. Results

Data Analysis

The survey results were read closely and categorized by different topics including their perception of assessment, the types of assessment used in their classrooms, and who is in charge of developing the assessment modes. Next, the interviews were used to further develop the themes that emerged from the survey. The interviews were coded by the first author and reviewed by the co-author multiple times. In more than one case, participants were reached out to in order to check and verify the overall assessment structure of the STEM schools and the grading system [48,55]. Data analysis of both sources led to the emergence of themes that are aligned with the research questions.

The results are presented as they pertain to each of the three research questions.

RQ1: What are the different assessment methods used in STEM schools in Egypt?

The survey results indicated that two broad types of assessment were used: (i) disciplinary assessment and (ii) interdisciplinary assessment. One participant succinctly described the differences in disciplinary and interdisciplinary assessments, "in the capstone, [assessment techniques] focus on the connection between subject areas, the quality of the product, the relevance to the project goals, and the troubleshooting (problem solving) through following the engineering design process (EDP), while the disciplinary assessments focus on the thinking patterns and skills".

Disciplinary assessments. Teachers described different classroom assessment tools and methods, including summative and formative assessments (see Table 4). The rationale they provided for using the various assessment tools varied, including checking understanding of the subject area knowledge and how this content knowledge can be integrated into the capstone. For example, Ebtisam explained that assessment is for "checking understanding and providing evidence of learning as well and to diagnose where students are and how I should be leading them". Other responses indicated that these tools were used to address different learning styles. Kareem and Omar described the process of planning assessment and instruction in STEM schools based on the understanding by design (UbD) framework [56]. Kareem maintained that "before the school year starts, we meet as colleagues and design some performance tasks for each subject that show mastery of the learning outcomes and their link to the capstone". Hadeer and Ebtisam described that they "use a variety of formative assessment methods to make sure that the learning outcome is mastered. The plan is most ideal but implementation-wise there should be alternative plans to accommodate students' conditions/levels". Ebtisam explained that "based on our experience with students and the challenges we have faced all over the years, I may change the sequence, the time, mode: whether it should be individual, pair or group".

The interviews further supported the use of formative and alternative assessment techniques described in Table 4. For example, Hadeer described how she integrated assessment into the learning experience from the outset of the lesson, starting with a "review to link the past learning experience to the present topic" using a KWL technique. Dalila described "pause and answer" in reading sessions as one of the most effective classroom techniques she has used to check for understanding. The use of internet-based assessment techniques was also evident. For example, Omar described using "videos on Youtube with questions to answer and Kahoot as an exit ticket".

The capstone as an interdisciplinary assessment. The predominant interdisciplinary assessment was the capstone project; Kareem and Ebtisam described it as a term-long project that allows students to collaboratively work on solving a real-life challenge to Egypt following the engineering design process (EDP) that starts with problem definition, research

solutions, identifying design requirements, identifying possible solutions, planning and designing the solutions, test and redesign, and finally communicating this solution.

Table 4. Assessment tools frequently mentioned in the survey ordered from most frequent to least.

Classroom Assessment Tools	Type
Brainstorming	Formative
Presentations	Formative
Class discussions	Formative
Quizzes	Summative at the of a learning unit
Inquiry based assessment	Formative/summative
Think pair share	Formative
Mini project-based assessment	Formative
Individual and group Hands on activities	Formative
KWL (what you Know–what you Want to know–what you have Learned)	Formative
Gallery walks	Formative
Exit ticket	Summative/Formative
Lab investigation	Formative/Summative
Jigsaw activities	Formative
Students leading class	Formative
Short written exercises	Formative/summative
Students debate	Formative
Reaction discussion/paper to a video	Formative
Frayer model	Formative/Summative
Games	Formative
Songs	Formative
Final exams	
At the disciplinary level	At the integrated level
University readiness test (URT) (evaluated centrally)	Capstone project (evaluated by external judges)
Concept inventory test (CT) (evaluated centrally)	

This process is a uniform process that all students follow in all of the STEM schools and is facilitated by a capstone facilitator (teacher) and a capstone coordinator. Teachers described both formative and summative assessments related to the capstone project. Based on the survey results, teachers described using journals and portfolios as formative assessments, a which were graded by the capstone teachers based on a rubric prepared by STEM experts overseeing the schools and affiliated to the MoE&TE STEM unit. The journal questions checked students' progress in terms of the skills being developed and used, how they overcome challenges they encounter, and the description of the engineering design stages at any given point in the project. As part of the journaling process, students also completed biweekly "transfer quizzes" designed to assess students' use of content knowledge and skills from different subjects in addressing the specific in their capstone projects. Transfer questions are questions targeting some subject topics that directly or indirectly help students with their work in the capstone project. These transfer questions were also prepared by external STEM experts with assistance from the teachers at the school. The questions are considered part of the overall grade for the capstone across all grade levels. The initial intent was for the transfer questions to be formative and provide some feedback to students, but they are no longer used to provide direct feedback to students. Students uploaded their answers, and the capstone leader distributed them to teachers for grading based on the rubrics provided.

The capstone grades are broken down according to the MoE&TE's decrees and regulations (see Table 5); when projects are finalized and students are ready with their posters and prototypes, an external committee evaluates the projects using the provided rubrics. The external committee evaluates the poster and the prototype, but the journals are evaluated by the teachers. The rubrics used in assessing the capstone are analytic in nature. These rubrics (see Appendix C) delineate expected performance from students regarding the

knowledge, skills, and attitudes students should have mastered. For instance, the section on the poster gives details of how the different components of the poster, abstract, introduction, results, discussion, and references should be evaluated on four levels: distinguished, accomplished, developing, pre-novice. Another section is devoted to the testable prototype where students are required to describe design requirements chosen for their prototype in a logical and well-reasoned manner. The prototype should be suitable for testing, and in the case of a software prototype (simulation) being used, the selection of modeling methods is logical and justified and students should be able to identify the functions or relationships contained in their software prototype.

Table 5. The cross-disciplinary modes of assessment (capstone).

Breakdown of the Different Assessment Modes in the Capstone Projects			
Formative and Ongoing	Weight	Summative	Weight
Journaling (include reflection questions on the process and transfer questions)	40%	Poster evaluation	40%
Portfolio	10%	Prototype evaluation	10%

RQ2: In what way, if any, do the teachers describe assessment systems used in STEM schools as different from the mainstream assessment systems?

Almost all STEM school teachers had prior teaching experience in non-STEM schools, and they were able to compare assessment use in the two systems. Responses from the survey indicated that the use of ongoing formative assessment and the capstone assessment are unique to STEM schools. One survey participant commented that mainstream schools rarely used assessment tools based on real-life problem solving, “hands on work”, for a variety of reasons such as “lack of resources, professional readiness, large number of students in the class, no systemic allotment of the ongoing formative or hands on work”. Participants with experience in mainstream schools cited the “traditional paper and pencil as the most common assessment technique in both the during the year and final assessment in the mainstream education schools”.

Another participant added that “in the STEM school, students can participate in designing assessment methods like designing open ended questions and leading a class with different activities including assessment activities. These types of questions target understanding instead of recall”. The same participant explained that “STEM focuses on higher level thinking, but traditional schools focus on recall and depend on the teacher rather than student, more practical skills integrating technology (assessment technology tools) rather than traditional paper and pencil questions”.

These differences were further elucidated in the interviews. For example, Mona stated “assessment in STEM is completely different from mainstream system and it is unfair to compare both (meaning as both use different instructional models, it is not a good idea that students have the same exam)”. She went on to explain, “it is unfair to assess thanaweya amma (high stakes Egyptian Secondary Certificate Exam) and STEM in the same exam (they cannot take the same exam) because they learn in completely different ways”. However, the expectations of the STEM assessment system led to large numbers of students leaving STEM schools in the final year as the “mainstream [standardized assessment] is less challenging as it is based on retrieval”. As Hadeer noted, “students learned the skills from STEM schools in grades 10 and 11 and moved to the mainstream to easily get the highest grades”. The issue is that the rigor of the assessments in the STEM schools is far exceeds that of the traditional exams in the Egyptian Secondary Certificate which mainly focuses on retrieval, and ultimately, the higher education admission system only recognizes the number of points earned on the final assessment.

RQ3: How are teachers' STEM assessment practices aligned with what STEM advocates? What are the challenges that hinder this alignment?

Approximately 90% of the surveyed teachers saw a high degree of alignment between the vision of the STEM schools and the assessment methods used. Some examples of school visions shared by teachers include: "to equip the students with 21st century's skill to face the challenges"; "to improve the quality of learning and students"; "to work in cooperative groups and learn through the integration of the different subjects, to use scientific and strategic methods to solve various problems that they face; to apply what they learn in the different subjects to scientific and technological projects"; "to be creative and discover all that is new in the various fields, especially in the scientific and technological fields"; "to contribute in, face and solve the contemporary grand challenges of Egypt, to build up generations of proactive, aware, effective, open-minded and leading students who can creatively contribute to facing their local and national challenges and keeping up with national and international changes". However, all such aspirations were only visible in the capstone work described by the teachers, with most of the other assessment methods mentioned by the teachers (for example, see Table 4) falling short of achieving these stated visions.

As a core feature in STEM education to have an alignment between instructional practices and assessment methods, the interviewed teachers described assessment as organically related to instruction. At the disciplinary level, Dalila explained that assessment is "not just asking students to present information but checking/gauging their deep understanding of what I teach". All interviewed teachers mentioned that they used rubrics in both their disciplinary and interdisciplinary work; rubrics where detailed descriptions of the expected work from students to show mastery of the intended learning outcomes help rigorous assessment of students' learning. They are considered an effective tool for ensuring a certain degree of alignment between intended outcomes and real achievement. The teachers referred to ideas that further intertwine instruction with assessment in a student-centered environment including activities based on flipped classroom technique whereas Hadeer described, "students are asked to prepare a topic at home; read assigned material, watch videos; and come to class ready to present to their peers what they have understood". Dalila maintained that "most of the class work is presentations. I start by asking driving or essential questions at the beginning of the session to drive the learning process in the class. Essential questions are used to guide the students' work in the session to facilitate their discussions and sharing ideas. We also use questions so that students do research, and debate". Teachers also described alignment between their individual subjects and the capstone. For example, Ebisam argued that one way of linking what she teaches in biology to the capstone on transportation was "modelling": she asked the students to create a model to test the potential in the nervous system based on the Nernst potential law. Ion channels conduct most of the flow of simple ions in and out of cells. Ions concentrate and the chemical gradient controls the potential of the membrane and hence, creates some form of the transportation of the cells.

Students were trying to show how traffic flow can be similarly modelled.

One of the major arguments for STEM education is building and consolidating 21st century skills (Authors, 2021) where students perform almost 60% of their capstone project tasks collaboratively, solve real world problems, and reflect on their learning. The rubrics used to evaluate these skills provide clear scales and criteria for assessing such skills. For example, the capstone projects are evaluated according to a rubric that measures the quality of (i) a poster which is the main communication medium of the whole project that includes an abstract, introduction, methods and materials, data, discussion etc.; (ii) construction of a testable prototype which is the hands-on product that presumably solve the problem at hand; (iii) data collection and analysis; and (vi) a capstone portfolio where is the journey of working on the project is recorded. Similarly, but to a lesser degree, the different disciplinary performance tasks are accompanied by such rubrics. For instance, the presentations students conduct in class, seen as a recurrent tool for formative assess-

ment of students' pre-class preparation, is assessed based on a rubric that assesses content, visuals, presentation skills, and collaboration, as in most cases these are performed in a group presentation format. In science subjects where laboratories are essential, there are different tools to assess students' ability to reason and support the use of evidence-based arguments. Discussions are facilitated by teachers to help build critical thinking skills. In this sense, based on the data from the survey and the interviews, the different modes, techniques, and tools of assessment used by the surveyed teachers provides a picture of how these methods reflect what STEM advocates: targeting 21st century skills, reliance on performance tasks, problem solving, and project-based assessments. However, full alignment between desired and actual assessment practices had several challenges.

Challenges.

More Autonomy. Ebtisam, Hadeer, and Kareem argued that teachers need to have more autonomy in designing and implementing STEM assessments. Teachers want an increased role in designing the biweekly journal reflection questions that are part of the capstone process, midterm disciplinary assessments that are centrally developed, and to participate in designing the questions of the final exams which are decreed to be designed by a MoE&TE committee. Ebtisam maintained that "we used to have a role when schools started but now very little is assigned to us. Teachers need to have a say in developing suitable exams for STEM students". Hadeer and Kareem echoed the same concern.

Almost all interviewed teachers, Kareem, Ebtisam, Omar, and Hadeer, mentioned that alignment was hindered as they "have never had access to any URT or CT exams the students have had". They explained that having these models can help us "reduce stress and anxiety for students and help teachers use more appropriate instructional practices". As noted by Kareem, the teachers assumed that the decision not to release the exams may be a way to "avoid making it a subject for private tuition and retrieval or teaching to the test".

Rigor. Another challenge is that rigor is compromised by not updating capstone assessments. Ebtisam and Kareem explained, "keeping the assessment tasks for some aspects of the capstone without being updated for a long time; these include the projects' challenges and the journal reflection prompts or questions". Since the initiation of the capstone project, the problems students solve, and the journal reflection questions/prompts have been the same which make it easy for students to copy from previous cohorts and thus the capstone assessment may not reflect the skills they were designed to assess.

In other words, what seems progressive can be misleading; a great tool of reflection like the biweekly journal may be used as a retrieval tool if the students know what the prompt is and even the answer itself from previous cohorts. Working on updating the tasks is a must. In addition, Ebtisam explained that in science subjects, "the biology practical exams are still the same ones we developed a long time ago when the STEM project started". As Ebtisam further noted, the different tasks designed for "the high stake items in the exams (like capstone challenges, transfer questions, reflection journal questions) should be continually updated so that students do not replicate the work and probably copy the answers of the previous cohorts". While the STEM vision calls for creativity and critical thinking, the reality of the assessment as a repetitive and predictable task creates a misalignment between the goals and reality of the capstone assessments. This is also evident in the way final assessments are conducted.

Final summative exams and capstone judges. The summative assessments, especially for during the final year of high school, are critical to students' entry to university, and teachers indicated two challenges related to the capstone and the URT final exam. Ebtisam and Kareem, capstone coordinators in their respective schools, expressed doubt about the readiness and efficiency of those who evaluate students' capstone work; "we need to make sure that the judges are well ready to do the evaluation: not superficial or so complex". One of the issues that need to be addressed is how an individual student can be assessed in a group project like the capstone; only the journal reflections and transfer questions give individual students an opportunity for being assessed on individual basis. The wide range of assessment methods used in formative assessment help assess individual students in

a group, but how are the contributions of individual learners assessed in the summative assessment of posters and prototypes?

At the URT exam level, which is hard for teachers to see, Ebtisam expresses her dismay as “some final summative exams continue to replicate the problems of the other standardized exams: they are blocking or threatening. Sometime we get to know the content of the exams from students as they leave the exam room. The summative assessment is still traditional; too high or out of scope just to challenge students. sometimes not related to the “learning outcomes” which hinders the alignment sought”. Hadeer sums it up, denoting that “based on the only trial or mock copy teachers have access to, the exam is a copy of so similar to ICT standardized exams”.

6. Discussion

The data from the survey and the semi-structured interviews lay out a unique picture of assessment in STEM schools that consists of both disciplinary and interdisciplinary assessments [36]. While teachers described the assessment provided in their STEM schools as a different learning experience compared to that provided in the mainstream system where students mostly undergo traditional retrieval classroom or standardized assessment [36,37]. However, this difference was limited to the capstone assessment. With the exception of the capstone which presented the interdisciplinary face of assessment in the STEM schools, it was difficult to see alignment between the vision and reality of STEM assessments. The standardized final exams, URT and CT, which contribute to 60% of students’ overall grade in the final year and 40% in the first and second year of the Egyptian STEM schools, are similar in nature to the standardized ICT exams within the mainstream system. Therefore, some kind of discrepancy was observed between what the majority of teachers initially maintained was an alignment between STEM schools’ vision or what STEM as an educational approach should provide, and the assessment practices they really use in their classes at both the disciplinary and interdisciplinary level, reflecting the mono-disciplinary–multidisciplinary dichotomy [36].

In their classrooms, where they have relative autonomy over assessment, teachers shared how assessment and learning processes are integrated in a complementary relationship [3,4]. This relationship relied on disciplinary-based assessments using formative assessment strategies that allowed teachers to assess conceptual understanding rather than memorization of facts, as well as providing important information for teachers to adapt their lesson plans [30]. However, while the disciplinary assessment system used a wide range of assessment methods, these assessments were not aligned with the stated goals and vision of the STEM schools which is a critical consideration given that constructive alignment is necessary for effective learning [12,13,52,53].

STEM schools rely on the capstone project and related assessments to address the goals and vision of STEM education in Egypt. The capstone project is also aligned with conceptual frameworks of STEM education that proposes project-based learning, problem solving, hands on, and design-based learning as its main instructional pillars [9,22]. However, in terms of assessment, more work is still needed to make authentic assessment a reality [32–34]; this can be achieved through ongoing design and redesigning of questions based on the students’ performance, developing new capstone challenges and prompts for the various capstone aspects (journals, transfer questions). Though the skills emphasized in the capstone tacitly target 21st century skills, assessing these skills is still not visible. These skills need to be explicit in the capstone rubrics in clear terms like stating “communication” as a component of poster and oral presentation assessment; collaboration as an essential part of the project; critical thinking and problem solving as the core of the Engineering Design Process; and creativity which is seen through the different products students come up with to solve the challenges they work on throughout their school journey.

Building consistency between the instructional pillars of STEM education and the assessment models/tools used requires effort from both teachers and curriculum developers [5,6]. The different models of assessment described by the teachers, especially in the

capstone, provide partial alignment with the goals of STEM education, if implemented as intended. From a bird's-eye point of view, an alignment can be seen [10–13]. However, a closer look reveals a weak alignment in action; the increasing reliance on standardized tests and final grade examinations; a reduction in the number of capstone projects to one project per year from two in grade 10 and 11 [53]; at the disciplinary levels, teachers working in isolation rather than integrated manner; and finally, the 21st century skills implicitly assessed but not clearly stated in the rubrics.

In this study, the specific experience of STEM education in Egyptian STEM schools denotes two parallel paths of assessment: one specific to a disciplinary competency or to a content area or domain [36,45], and one designed especially for competencies that are attributed to more than one STEM discipline such as designing a solution in an engineering design context and develop 21st century skills; two big promises of STEM education [22,43]. The greatest challenge, though, is to create an organic and cohesive relationship between both aspects that will result in an effective assessment system of STEM education process where expectations and reality match.

7. Limitations and Further Research

One of the limitations of this study is the lack of direct observation of teachers conducting their classroom assessment to check the real practices teachers do inside their classes. A future follow-up of this study will take this in consideration.

8. Conclusions

The reality of assessment in Egyptian STEM setting was explored to check to what extent it is aligned with the expectations of STEM education as set in research and policy documents. Effective quality assessment in STEM schools moves back and forth from the siloed to the interdisciplinary level, leaving a weak or partial alignment with STEM schools' visions and STEM approach aspirations. For example, a question or a task prompt that is used over and over for successive cohorts, will turn into a traditional rote learning and retrieval task; no matter how deep and well-structured it is. In the ideal case scenario these assessment methods need to be clearly written, well-structured, and reflective of the objectives of both the disciplinary and interdisciplinary levels of STEM. For this to happen, teachers need more autonomy, trust, and resources to develop their own assessment tools to avoid the dichotomy between what is intended and what is really done.

Author Contributions: M.A.E.N.: Conceptualization, Methodology, Data curation and analysis, writing original draft and the final manuscript. G.H.R.: Validation and investigation, review and editing; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of American University in Cairo, "CASE #2020-2021-013" on 15 October 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: Authors acknowledge the support of Heba El-Deghaidy h.eldeghaidy@aucegypt.edu for reviewing an initial draft of this article; and Benny Mart Hiwatig hiwat001@umn.edu for initial literature review.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Here are the survey questions:

1. What subject do you teach?

2. How long have you been teaching in your (STEM) school?
3. How long have you been teaching overall?
4. Roughly, how many students are there in your school?
5. What is your school mission and/or vision?
6. How often do you include integrated STEM activities in your classes? Please list the STEM activities that you implement.
7. What kind of assessment do you use in your (STEM) school at your disciplinary level: science, math, English, social studies?
8. What kind of assessment do you use in your STEM school at the capstone/integrated project level, if any?
9. Please share a sample learning outcome and a sample assessment you use to assess that learning outcome.
10. Are you the person in charge of designing your own assessments? If not, who and why? Explain if this differs according to grade level.
11. How different are your assessments in the STEM school to those used in your previous non-STEM schools?
12. How are your assessments for STEM different to assessments using a disciplinary approach (e.g., science or mathematics)?

Appendix B

Here are the interview questions:

1. In what ways do you implement STEM lesson plans in your classroom?
2. What are your goals for students when using STEM in your classroom?
3. What does assessment mean for you?
4. How do you assess student learning in a STEM lesson or unit?
 - a. Use follow-up questions if they do not address content learning, teamwork, engineering design
 - b. Follow-up to find out if they use rubrics
 - c. Follow-up to find out if they use STEM or engineering notebooks and how these are assessed
5. In what ways is assessment in STEM different to science?
6. In what ways is STEM assessed in standardized state or national testing?
7. What is the most common type of assessment do you use in your STEM setting?
8. Follow up question: Which do you believe is the most effective type of these assessments in your STEM setting?
9. What is the difference between these forms of assessment you use and those used in the mainstream non-STEM education systems?

Appendix C. Poster Level of Proficiency

Criteria	Distinguished	Accomplished	Developing	Pre-Novice
Abstract (Poster) %5	Distinguished includes all of the “Accomplished” criteria and the following. ___Abstract clearly ties together the entire project from Grand Challenge to chosen solution, design requirements and prototype, testing results and conclusions. ___Abstract alone generates excitement and desire to learn more about the topic. ___Writing is professional, organized and well developed.	___Abstract is a brief description of the entire work described in the poster. ___Abstract understandable without reading the entire poster. ___Includes (1) purpose of the study, (2) brief statement of what was done (without including minor details of the methods), (3) brief statement of major findings, and (4) major conclusions. ___Writing is clear and readable.	___Abstract present and relatively complete but not prepared according to all guidelines.	___No abstract.
Introduction (Poster) %20	Distinguished includes all of the “Accomplished” criteria and the following. ___Moves clearly from a broad view of the Grand Challenge and research on various solutions to an increasingly narrow focus on the team’s chosen solution, design requirements, and prototype, justifying each choice. ___Provides a smooth transition from “what” choices they made and “why” they made them to the upcoming Materials/Methods section (the “how” section of the poster).	___Connection is made to Egypt’s Grand Challenges. ___Clearly and objectively identifies the problem and summarizes prior solution attempts strengths and weaknesses. ___Includes design requirements for a new solution that can be tested. ___Summarizes how the team’s solution was chosen and how it addresses design requirements.	___Introduction present and relatively complete but does not address all points indicated.	___No introduction.
Materials and Methods (Poster) %10	Distinguished includes all of the “Accomplished” criteria and the following. ___The methods are clear enough to permit a reader to explain the method to another professional and be able to replicate the method.	___A summary of the test plan for the prototype includes a summary of tests conducted and how they address design requirements. ___Materials lists and/or illustrations are summarized.	___Test Plan Methods and Materials lists present and relatively complete but does not address all points indicated.	___No Methods or no Materials list.
Results (Poster) %15	Distinguished includes all of the “Accomplished” criteria and the following. ___Supporting documentation (Capstone Portfolio) contains all data collected and is so well organized it could be handed to a new team to replicate the work with high fidelity. ___The visual representation of the results alone (without the words in the Results section) leads the reader to a conclusion about the results.	___All types of results are presented, whether positive or negative. ___The Capstone Portfolio is available to show data for tests or scenarios that were conducted. ___Includes a table or figure that is appropriate for the type of results being described.	___Results present and relatively complete but does not address all points indicated.	___No results.

Discussion (Poster) %35	Distinguished includes all of the “Accomplished” criteria and the following. ___Conclusions are drawn from the test results and then compared with the team’s research on other solutions. ___Recommendations are practical and directed towards a future research, engineering or policy group. Recommendations are clearly informed by the problem, their proposed solution, and their findings. ___Students can articulate specific evidence of learning transfer from two or more of their content classes (learning outcomes) this Semester into their Capstone.	___Discussion ties performance results to the original question being addressed and to the Grand Challenge. ___Proposed solution is supported with robust STEM principles and demonstrates applied learning transfer. ___Analysis is supported by pictures, graphs, charts and other visuals, and test results. ___Recommendations for future study are provided, including specific ways the project could be improved in the future. ___Writing is clear, organized and well developed. It explains, questions or persuades. It is written to meet the needs of the intended audience, and it uses forms that are common among STEM disciplines (E.g., notes, descriptive/narrative accounts, research reports).	___Discussion present and relatively complete but does not address all points indicated.	___No discussion.
Literature Cited (Poster) %5	Distinguished includes all of the “Accomplished” criteria and the following: ___At least five citations are peer-review publications.	___Includes only sources cited in the poster text (at least 5 sources). ___Includes only papers actually read by the students. ___Is prepared according to the American Psychological Association (APA) style guidelines.	___Literature cited present and relatively complete but does not address all points indicated.	___No literature cited (appropriate only if no citations used in the text).
(Poster) Title, Name, Affiliation, Size, Layout, Graphics, Tables, Photos, Other Images %10	Distinguished includes all of the “Accomplished” criteria and the following. ___The poster demonstrates brevity (focused, well synthesized and straight to the point) while targeting a professional audience. ___Visuals are well titled and labeled and tell clear stories with no other supporting text necessary.	___Text is readable from a distance of about 1 m. ___Title is at top of the poster, short, descriptive of the project and easily readable at a distance of about 2 m (words about 1.5–2.5 cm tall). ___Includes presenter’s name and school’s name in a section about 20–30% smaller than the title. ___Illustrations, tables, figures, photographs or diagrams have unique identification numbers and a key to identify symbols. ___Text includes references to specific graphics or pictures ___Legends include full explanation and where appropriate, color keys, scale, etc. ___All images presented in appropriate layout and size relative to text.	___Elements are present but not meeting all of the requirements in the “Accomplished” column, or some elements are missing.	___No graphics, tables, photos or other images (the poster should include at least some images).

Criteria	Prototype Level of Proficiency			
	Distinguished	Accomplished	Developing	Pre-Novice
Construction of a testable prototype (Prototype) %50	Distinguished includes all of the “Accomplished” criteria and the following. ___Prototype is directly relevant to the chosen solution (E.g., an actual water treatment step). ___Students can demonstrate the functionality of their prototype or show visual proof of functionality in a different environment (e.g., Lab). ___If a software prototype (simulation) is used, modeling software such as LabView, Excel, or a programming language is demonstrated and can be tested.	___Students can describe design requirements chosen for this prototype. Choice of design requirements is logical and well-reasoned. ___A prototype has been constructed that was suitable for testing. ___If a software prototype (simulation) was used, the selection of modeling methods is logical and justified and students can identify the functions or relationships contained in their software prototype.	___A prototype or model has been constructed, some justification for the selection of design requirements and modeling approach is provided, but it is incomplete.	___No prototype or model; or no evidence that constructed prototype or model would facilitate test of any of the design requirements.
Capstone Portfolio: Prototype Testing and Data Collection Plan (Prototype) %20	Distinguished includes all of the “Accomplished” criteria and the following. ___Capstone Portfolio contains a test plan that clearly connects every type of test listed to a specific design requirement, and all chosen design requirements are addressed by the test plan. ___Capstone Portfolio clearly communicates how each test in the test plan isolates what is being tested and acknowledges other interfering factors that might affect the results. ___Capstone Portfolio contains measurement methods that have quantified error described explicitly (e.g., $+/-0.1$ Volts). ___Capstone Portfolio indicates that the total materials expenditures are within budget (evaluated by administration).	___Capstone Portfolio contains a test plan for the prototype which provides the scenarios to be tested and how they relate to design requirements. ___Capstone Portfolio contains a test plan that describes tests to be conducted in a thorough and clearly understandable manner. ___Capstone Portfolio test plan supports repetition and testing by others. ___Capstone Portfolio indicates that the total materials expenditures are below 1.5 times the budget (evaluated by administration).	___Testing plan exists and partially describes the testing to be conducted; limited justification of why the tests were selected.	___Testing plan is missing altogether, or fails to demonstrate any understanding of why tests relate to the design requirements.

Capstone Portfolio: Testing, data collection and analysis (Prototype) %30	Distinguished includes all of the “Accomplished” criteria and the following. ___The Materials list is clear enough that a reader can replicate the work precisely. ___Capstone Portfolio accurately records all data collected through the project phases. ___Capstone Portfolio is sufficiently well organized that it can be turned over to another group to replicate the prototype and tests. ___Capstone Portfolio lists 10 learning outcomes from their other subjects that they have transferred and applied in their capstone. Each learning outcome must have one paragraph clearly explaining how this learning outcome was transferred to their Capstone project.	___Capstone Portfolio contains a material list with cost (includes receipts if purchased) and/or illustrations, needed to replicate the prototype or model. ___Capstone Portfolio demonstrates whether the prototype met design requirements with data from each portion of the test procedure ___Capstone Portfolio contains analysis supported by graphs, charts and/or other visuals. ___Capstone Portfolio lists 5 learning outcomes from their other subjects that they have transferred and applied in their capstone. Each learning outcome should have one paragraph explaining how this learning outcome was transferred to their capstone.	___Documentation is provided for some tests that were conducted, but some are not described. ___Analysis of the effectiveness of the design is generally described and has limited support using pictures, graphs, charts and other visuals.	___No documentation presented for test results, or results presented are not tied to the testing plan in any logical way.
---	---	--	--	--

References

- Freeman, B.; Marginson, S.; Tytler, R. The age of STEM: Educational policy and practice across the world in science, technology, engineering and mathematics (1st ed.). *Taylor Fr.* **2014**, *26*, 303. [CrossRef]
- Cunningham, C. *Engineering in Elementary STEM Education: Curriculum Design, Instruction, Learning, and Assessment*; Teachers College Press: New York, NY, USA, 2018; ISBN 13:978-0807758779.
- Faxon-Mills, S.; Hamilton, L.S.; Rudnick, M.; Stecher, B.M. *New Assessments, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement*; RAND Corporation: Santa Monica, CA, USA, 2013.
- Kinash, S.; Knight, D. *Assessment @ Bond*; Office of Learning and Teaching, Bond University: Gold Coast, QLD, Australia, 2013; ISBN 9781922183118.
- Roller, S.A.; Cunningham, E.P.; Marin, K.A. Photographs and Learning Progressions. *YC Young Child.* **2019**, *74*, 26–33.
- Sato, M.; Wei, R.C.; Darling-Hammond, L. Improving teachers’ assessment practices through professional development: The case of national board certification. *Am. Educ. Res. J.* **2008**, *45*, 669–700. [CrossRef]
- Baird, J.; Andrich, D.; Hopfenbeck, T.N.; Stobart, G. Assessment and learning: Fields apart? *Assessment in Education. Princ. Policy Pract.* **2017**, *24*, 317–350.
- Johnson, C.C.; Moore, T.J.; Utley, J.; Breiner, J.; Burton, S.R.; Peters-Burton, E.E.; Walton, J.B. The STEM road map for grades 6–8. In *STEM Road Map 2.0*; Routledge: London, UK, 2021; pp. 102–132.
- National Research Council. *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research*; National Academies Press: Washington, DC, USA, 2014.
- Biggs, J. *Teaching for Quality Learning at University—What the Student Does*, 2nd ed.; Open University Press: London, UK, 2003.
- Biggs, J.B.; Tang, C.K. *Teaching for Quality Learning at University: What the Student Does*, 4th ed.; Open University Press: London, UK, 2011.
- Martone, A.; Sireci, S. Evaluating Alignment between Curriculum, Assessment, and Instruction. *Rev. Educ. Res.* **2009**, *79*, 1332–1361. Available online: <http://rer.aera.net> (accessed on 1 August 2022). [CrossRef]
- McMahon, T. Achieving constructive alignment: Putting outcomes first. *Aukštojo Moksl. Kokyb.* **2006**, *3*, 10–19.
- Borrego, M. Constructive alignment of interdisciplinary graduate curriculum in engineering and science: An analysis of successful IGERT proposals. *J. Eng. Educ.* **2010**, *99*, 355–369. [CrossRef]
- Chase, A. A report of the STEM education track of the 2017 assessment institut. *Assess. Update* **2018**, *30*, 6–7. [CrossRef]
- El Nagdi, M.; Roehrig, G.H. Gender equity in STEM education: The case of an Egyptian girls’ school. In *Theorizing STEM Education in the 21st Century*; IntechOpen Publications: London, UK, 2019; pp. 315–317. Available online: <https://www.intechopen.com/chapters/67951> (accessed on 1 April 2022).
- National Research Council. *Monitoring Progress toward Successful K-12 STEM Education: A Nation Advancing?* National Academies Press: Washington, DC, USA, 2013.

18. Perignat, E.; Katz-Buonincontro, J. STEAM in practice and research: An integrative literature review. *Think. Ski. Creat.* **2019**, *31*, 31–43. [CrossRef]
19. Bybee, R.W. Advancing STEM Education: A 2020 Vision. *Technol. Eng. Teach.* **2010**, *70*, 30.
20. Bybee, R.W. *The Case for STEM Education: Challenges and Opportunities*; NSTA press: Arlington, TX, USA, 2013.
21. Zollman, A. Learning for STEM literacy: STEM literacy for learning. *Sch. Sci. Math.* **2012**, *112*, 12–19. [CrossRef]
22. Roehrig, G.H.; Dare, E.A.; Ellis, J.A.; Ring-Whalen, E. Beyond the basics: A detailed conceptual framework of integrated STEM. *Discip. Interdiscip. Sci. Educ. Res.* **2021**, *3*, 1–18. [CrossRef]
23. Goldstein, H. A response to assessment and learning: Fields apart? *Assess. Educ. Princ. Policy Pract. Assess. Learn.* **2017**, *24*, 388–393. [CrossRef]
24. Ring, E.A.; Dare, E.A.; Crotty, E.A.; Roehrig, G.H. The evolution of teacher conceptions of STEM education throughout an intensive professional development experience. *J. Sci. Teach. Educ.* **2017**, *28*, 444–467. [CrossRef]
25. Breiner, J.M.; Harkness, S.S.; Johnson, C.C.; Koehler, C.M. What is STEM? A discussion about conceptions of STEM in education and partnerships. *Sch. Sci. Math.* **2012**, *112*, 3–11. [CrossRef]
26. Herschbach, D.R. The STEM initiative: Constraints and challenges. *J. STEM Teach. Educ.* **2011**, *48*, 96–122. [CrossRef]
27. El Nagdi, M.; Roehrig, G. Identity evolution of STEM teachers in Egyptian STEM schools in a time of transition: A case study. *Int. J. STEM Educ.* **2020**, *7*, 1–16. [CrossRef]
28. Roehrig, G.; El-Deghaidy, H.; García-Holgado, A.; Kansan, D. A closer look to STEM education across continents: Insights from a multicultural panel discussion. In Proceedings of the 2022 IEEE Global Engineering Education Conference (EDUCON), Tunis, Tunisia, 28–31 March 2022; pp. 1873–1880.
29. El-Deghaidy, H. STEAM Methods. In *Designing and Teaching the Secondary Science Methods Course*; Brill: Leiden, The Netherlands, 2017; pp. 71–87.
30. Stiggins, R. From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan* **2005**, *87*, 324–328. [CrossRef]
31. Westbroek, H.B.; van Rens, L.; van den Berg, E.; Janssen, F. A practical approach to assessment for learning and differentiated instruction. *Int. J. Sci. Educ.* **2020**, *42*, 955–976. [CrossRef]
32. Palm, T. Performance assessment and authentic assessment: A conceptual analysis of the literature. *Pract. Assess. Res. Eval.* **2008**, *13*, 4. Available online: <https://scholarworks.umass.edu/pare/vol13/iss1/4> (accessed on 15 August 2021). [CrossRef]
33. Villarroel, V.; Bloxham, S.; Bruna, D.; Bruna, C.; Herrera-Seda, C. Authentic assessment: Creating a blueprint for course design. *Assess. Eval. High. Educ.* **2018**, *43*, 840–854. [CrossRef]
34. Wiggins, G. The case for authentic assessment. *Pract. Assess. Res. Eval.* **1990**, *2*, 1–3.
35. Tan, A.L.; Leong, W.F. Mapping Curriculum Innovation in STEM Schools to Assessment Requirements: Tensions and Dilemmas. *Theory Pract.* **2014**, *53*, 11–17. [CrossRef]
36. Gao, X.; Peishan, L.; Shen, J.; Sun, H. Reviewing assessment of student learning in interdisciplinary STEM education. *Int. J. STEM Educ.* **2020**, *7*, 1–14. [CrossRef]
37. Harwell, M.; Moreno, M.; Phillips, A.; Guzey, S.S.; Moore, T.J.; Roehrig, G.H. A Study of STEM Assessments in Engineering, Science, and Mathematics for Elementary and Middle School Students. *Sch. Sci. Math.* **2015**, *115*, 66–74. [CrossRef]
38. Hansen, M.; Gonzalez, T. Investigating the relationship between STEM learning principles and student achievement in math and science. *Am. J. Educ.* **2014**, *120*, 139–171. [CrossRef]
39. Ing, M. Can parents influence children's mathematics achievement and persistence in STEM careers? *J. Career Dev.* **2014**, *41*, 87–103. [CrossRef]
40. Seage, S.J.; Türegün, M. The Effects of Blended Learning on STEM Achievement of Elementary School Students. *Int. J. Res. Educ. Sci.* **2020**, *6*, 133–140. [CrossRef]
41. Van der Vleuten, C.P. Assessment in the context of problem-based learning. *Adv. Health Sci. Educ. Theory Pract.* **2019**, *24*, 903–914. [CrossRef]
42. Bicer, A.; Capraro, R.M.; Capraro, M.M. Integrated STEM assessment model. *Eurasia J. Math. Sci. Technol. Educ.* **2017**, *13*, 3959–3968. [CrossRef]
43. Moore, T.J.; Stohlmann, M.S.; Wang, H.H.; Tank, K.M.; Glancy, A.W.; Roehrig, G.H. Implementation and integration of engineering in K-12 STEM education. In *Engineering in Pre-College Settings: Synthesizing Research, Policy, and Practices*; Purdue University Press: West Lafayette, IN, USA, 2014; pp. 35–60.
44. Arikan, S.; Erktin, E.; Pesen, M. Development and validation of a STEM competencies assessment framework. *Int. J. Sci. Math. Educ.* **2020**, *20*, 1–24. [CrossRef]
45. Pellegrino, J.W. Proficiency in science: Assessment challenges and opportunities. *Science* **2013**, *340*, 320–323. [CrossRef] [PubMed]
46. Septiani, A.; Rustaman, N.Y. Implementation of performance assessment in STEM (Science, Technology, Engineering, Mathematics) education to detect science process skill. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2017; Volume 812, p. 012052.
47. Yin, R.K. *Case Study Research: Design and Methods*, 5th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2014.
48. Creswell, J. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th ed.; SAGE Publications: Thousand Oaks, CA, USA, 2014.

49. Greene, J.C.; Caracelli, V.J.; Graham, W.F. Toward a conceptual framework for mixed method evaluation designs. *Educ. Eval. Policy Anal.* **2016**, *11*, 255–275. [[CrossRef](#)]
50. U.S. Agency for International Development (USAID). Egypt STEM School Project (ESSP) FINAL REPORT. ESSP Final Report, USAID/Egypt Cooperative Agreement No. AID 263-A-12-00005. 2017. Available online: https://pdf.usaid.gov/pdf_docs/PA00THDB.pdf (accessed on 15 August 2021).
51. U.S. Agency for International Development (USAID). Support for STEM Secondary Education: Egypt. U.S. Agency for International Development. 2020. Available online: <https://www.usaid.gov/egypt/documents/support-stem-secondary-education> (accessed on 23 April 2022).
52. Ministerial Decree 382. Rules of Students' Admission, Study and Assessment for STEM High School. 2012. Available online: <http://moe.gov.eg/stem/12-382.pdf> (accessed on 15 August 2021).
53. Ministerial Decree 238. System of Examination and Certificate of Completion Awarded to STEM High School Student. 2013. Available online: <https://manshurat.org/node/2620> (accessed on 15 August 2021).
54. Rissmann-Joyce, S.; El Nagdi, M. A case study: Egypt's first STEM schools: Lessons learned. In Proceedings of the Global Summit on Education (GSE2013), Kuala Lumpur, Malaysia, 11–12 March 2013.
55. Saldaña, J. *The Coding Manual for Qualitative Researchers*, 2nd ed.; SAGE: Newcastle upon Tyne, UK, 2013.
56. Wiggins, G.; Wiggins, G.P.; McTighe, J. *Understanding by Design*; ASCD: Alexandria, VA, USA, 2005.