Deconstructing Physics Problems Using Testlet Templates

Budi Naini Mindyarto[1], Sugiyanto Sugiyanto[2]
{budinaini@mail.unnes.ac.id[1], sugiyanto@mail.unnes.ac.id[2]}

Universitas Negeri Semarang, Sekaran Gunungpati Semarang, Indonesia[1,2]

Abstract. Physics problems have been deconstructed into testlet items using templates. These items have been generated based on the functional requirements of a developed automatic item generator. Empirical testing needs to be conducted to verify the conformity with automatic item generation characteristics. This paper aims to address the extent of template compliance with the generator features. Black-box testing procedure was used to verify the function conformities by exploring the execution of the generator and its input and output. 25 Physics templates were examined. The testlet variants generated from the execution of the generator showed that the template characteristics adhered to the input requirements of the generator. It can be concluded that the validation of the testlet templates by the generator showed that the templates support the automatic item generation.

Keywords: physics problems, testlet templates, automatic item generation

1 Introduction

Traditional item development using a paper-and-pencil based manual process is not efficient. This is more pronounced when a large number of items are needed, such as to fulfill the demands of the items in parallel test packages or a question bank. Items are treated as isolated entities that are individually created, reviewed, and formatted. Because the items were developed individually, these items gave unpredictable statistical output because incidental and radical elements were not easily identified or understood. Automatic item generation (AIG) [1,2] can accommodate the deficiencies of traditional methods to generate items.

AIG is a process of using item models to generate test items with the help of computer technology. Computer-based algorithms are used to place the material into the item model through programming to generate items automatically. Without using any constraints, all the contents of the variables are combined systematically to generate items iteratively. Constraints are used to eliminate items that do not make sense or have no meaning.

The generation of items has been implemented for various purposes. Higgins, Futagi, and Deane [3] describe the ModelCreator for producing mathematical narrative questions, Higgins [4] presents ItemDistiller as a tool that can be used sentence-based items, Gierl et al. [5] describe the IGOR (Item GeneratOR) software to generate template-based items, Gutl et al. [6] describes the use of Enhanced Automatic Question Creator (EAQC) to retrieve key concepts from the text to generate multiple-choice items and constructed responses, Ferreyra & Backhoff Escudero [7] describe GenerEx to generate different versions of a basic

AECon 2020, December 19-20, Purwokerto, Indonesia

DOI 10.4108/eai.19-12-2020.2309165

competency test, Kosh [8] creates AIG item models to assess middle school students' algebraic reasoning, and Rambhau et al. [9] to generate MCQ with random answer key.

For objective items that present alternative answer choices where the testee is asked to choose one correct answer, the implementation of automatic item generation that has been done a lot is for the multiple-choice format. The response to this multiple-choice item is dichotomous. The design and implementation of the automatic item generation technique for the multiple-choice format have the potential to be modified to accommodate the generation of items for the testlet format that provides a polytomous response. The cognitive model will include the identification of the material to be tested, the development of cognitive processes that will be revealed, and the development of an item structure that covers the entire testlet. The author has developed a generator for producing testlet automatically.

The generator is an application program for generating testlet variants based on testlet models formatted in template forms. The program accepts and validates input in the form of a testlet template from the user, processes the testlet template into XML format, and generates testlet variants based on the testlet template and stored in the database. The testlet is a combination of multiple-choice items with both single correct answers (multiple-choice) and with multiple correct answers (multiple responses). The validation of the template‘s functionality is presented in Table 1.

Table 1. Fitures of Testlet Template

No.	Function Specifications
1	Manually input the testlet template or upload a template (.docx).
2	Image files for the template are uploaded separately.
3	Images can be used in stem and options.
4	Numeric expressions or formulas can be assigned to stem and options.
5	The key to the testlet question can be fixed or conditional.
6	The key to the testlet question can be single or multiple.

The author has developed physics testlet templates based on re-engineering the national exams of high school physics. Re-engineering is carried out with a motivation, namely stimulating the testee to carry out a qualitative analysis in solving quantitative physics problems. The stimulation is done by adding qualitative questions to the expanded context along with the reasoning questions. Re-engineering is carried out for problems concerning mechanics, namely for the topics of Kinematics and Dynamics, Elasticity and Hooke's Law, Work and Energy, the Law of Conservation of Mechanical Energy, and Impulses, Momentum and Collisions. An example of the re-engineering testlet template is presented in Figure 2.

Physics testlet templates developed by re-engineering high school physics national exams need to be validated and verified so that they can be used to generate physics testlet variants automatically. The validation of these physics templates is determined through black-box testing by exploring the execution and input-output of the generator. The success of the validation of these physics templates is indicated by the generation of physics testlet variants by the generator. So this research aims to explain qualitatively the fit between the characteristics of the physics testlet templates and the features of the generator.

The validation and verification are testing steps against the physics testlet templates which were developed by re-engineering the physics national-exam questions. The testing step is in line with Bohdan & Vasyl's [10] and Hamza & Hammad [11] statement that testing is an integral part of a development process that forms an important link in the overall development chain. It is expected that these physics testlet templates can be used appropriately with their function as templates in automatic item generation and it is hoped that these physics testlet templates can also show this function correctly. The suitability of the characteristics of the physics templates with the generator features can be used as a reference for developing testlet templates for other domains of science.


```
<stem>
    <idtestlet>MEKAN01</idtestlet>
    <idtopik>MEKAN</idtopik>
    <iduser>budinaini</iduser>
    <stemtestlet/>
    <filepicstemtestlet1/>
    <filepicstemtestlet2/>
    <nvar>4</nvar>
    <kd1>N1</kd1>
    <nil1>2|3</nil1>
    <kd2>N2</kd2>
    <nil2>1.25|1.8|3.2</nil2>
    <kd3>S1</kd3>
    <nil3>dijatuhkan bebas|meluncur pada bidang miring
tanpa gesekan| meluncur pada lintasan parabola tanpa
gesekan</nil3>
    <kd4>S3</kd4>
    <nil4>Pernyataan mendasar yang mendukung|Prinsip yang
mendasari|Argumen yang dapat menjelaskan</nil4>
    <nsyarat>0</nsyarat>
    <nvarkond>3</nvarkond>
    <eks1varkond1>S1</eks1varkond1>
    <eks2varkond1>dijatuhkan bebas</eks2varkond1>
    <kdvarkond1>S2</kdvarkond1>
    <nilvarkond1>mekanik0102a.png</nilvarkond1>
    <eks1varkond2>S1</eks1varkond2>
    <eks2varkond2>meluncur pada bidang miring tanpa
gesekan</eks2varkond2>
    <kdvarkond2>S2</kdvarkond2>
    <nilvarkond2>mekanik0102b.png</nilvarkond2>
    <eks1varkond3>S1</eks1varkond3>
    <eks2varkond3>meluncur pada lintasan parabola tanpa
gesekan</eks2varkond3>
    <kdvarkond3>S2</kdvarkond3>
    <nilvarkond3>mekanik0102c.png</nilvarkond3>
</stem>
```


Fig. 2. Conservation of Mechanical Energy Template

2 Method

The research design was descriptive qualitative related to developing physics testlet templates. The templates focused on developing knowledge structures [12] through-provoking deeper context-understanding of what-if questions [13]. The testlet accommodated the context expansion of the problems to reveal more comprehensive understanding. A what-if question and its reasoning are inserted into the testlet template to stimulate a deeper understanding of the problems. The implemented testlet template was visualized in Figure 3.

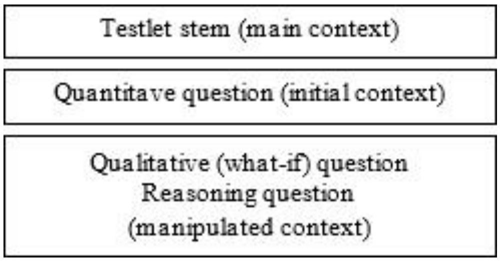


Fig. 3. Implem*en*ted template structure

The research description was obtained through functional testing or black-box testing (BBT), which verified the program's proper handling of external functions by observing program behavior during execution [10,11]. The program, which was the item generator, worked by combining the manipulative variables declared in the testlet templates to generate item variants. The form of BBT used was a checklist in the form of a functional specification that contained external functions that must be present and expected input-output information. The instrument facilitated the identification and verification of characteristic compatibility between the physics testlet templates and the generator in the perspective of automatic item generation. The procedure was to enter the templates into the generator, validate the templates, and generate testlet variants using the generator. The execution observation data was used to verify checklist items. This data was analyzed qualitatively to determine the fulfillment of templates with the generator features in the perspective of automatic item generation.

3 Results and Discussion

The data describing the functionality of physics testlet templates were obtained empirically by inserting testlet templates, validating, and generating testlet variants. The data were analyzed using a checklist analysis to match the characteristics of the physics testlet templates with the standard criteria, namely the generator features. The verification of the AIG perspective from these physics testlet templates was carried out based on the Table 1 criteria.

The generated items were automatically traced using manipulative variables in the physics testlet templates. The physics testlet variants were generated by the generator by combining all the values of all the manipulative variables used in the physics testlet template. For example, the selected testlet variants generated from the MEKAN01 testlet template and the values of the manipulative variables were presented in Table 2.

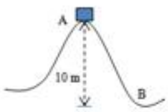
The MEKAN01 template used two manipulative variables (numeric: N1 and N2), two manipulative variables (string: S1, S3) and one conditional manipulative variable (image: S2). The values of the manipulative variables in Table 2 were substituted as a substitute for the manipulative variable codes that were in the MEKAN01 testlet template for each testlet variant. A snippet of the serial number 13 testlet variant generated from the MEKAN01 testlet template was presented in Figure 4.

Table 2. Variation of Manipulative Variables for Several Variants

Variant	Quantitative	Qualitative (what-if)	Scientific reasoning
13	N1=2; N2=1.8	S1 = ‘slide on a frictionless inclined plane’; S2 = ‘mekanik0102b.png’	S3 = ‘Supporting fundamental statements’
35	N1=3; N2=1.25	S1 = ‘slide on a parabolic trajectory without friction’; S2 = ‘mekanik0102c.png’	S3 = ‘Underlying principles’
48	N1=3; N2=3.2	S1 = ‘free dropped’; S2 = ‘mekanik0102a.png’	S3 = ‘Explanatory arguments’

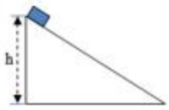
Problem 13

An object with mass of 2 kg moves from point A at an altitude of 10 meters without initial velocity. If during motion there is no friction, the mechanical energy of the object when its height decreases by 1.8 meters from point A is:



A. 3.6 joule
B. 36 joule
C. 72 joule
D. 164 joule
E. 328 joule

a) If the same object slides on an inclined plane without friction from the same height (h), then the mechanical energy of the block in the current lowest position is ...



A. becomes bigger
B. becomes smaller
C. fixed

b) The basic statement which supports the ratio of the mechanical energy of the block at the lowest position is:
A. The masses of the two blocks are equal.
B. The blocks move from the same initial height.
C. Both blocks move without initial velocity.
D. Shorter beam trajectory
E. The beam trajectory is longer.

Fig. 4. Testlet Variant Code 13 Generated from the Template MEKAN01

With five variables consisting of three manipulative variables and one conditional manipulative variables, the MEKAN01 template produced 54 testlet variants. The number of manipulative variables used per testlet template and the number of testlet variants generated by the generator from 25 physics testlet templates which were the objectives of this study were presented in Table 3.

Table 3. Distribution of Manipulative Variables and Variants per Template

Sub-Topics	Template Code (Σ variables | Σ variants)				
	T1	T2	T3	T4	T5
Kinematics and Dynamics	KINED-01 3|72	-02 4|72	-03 4|60	-04 5|90	-05 9|120
Elasticity and Hooke's Law	ELAST-01 10|432	-02 4|90	-03 4|32	-04 5|216	-05 9|108
Work and Energy	USAHA-01 6|360	-02 9|216	-03 4|90	-04 4|120	-05 4|120
Law of Conservation of Mechanical Energy	MEKAN-01 5|54	-02 9|384	-03 3|75	-04 12|72	-05 11|216
Impulse, Momentum and Collisions	TUMBU-01 9|312	-02 6|180	-03 6|216	-04 6|384	-05 6|168

The number of testlet variants generated from the testlet template depended on the number of manipulative variables in the template and the number of values or value ranges of each manipulative variable. The distribution of the testlet variants for the 25 physics testlet templates as presented in Table 3 showed that the physics testlet templates already had an automatic item generation character where the templates were successfully validated and processed to generate testlet variants. The success of processing testlet templates as input and the variants of physics testlets generated by the generator program as output showed that the physics testlet templates had characteristics that match the input requirements of the generator program for AIG. The template characteristics and the input requirements of the generator represented the properties of *creative task* and *generative task* [14] of AIG.

The validation facts of the 25 physics testlet templates developed by re-engineering the national-exam physics questions and generating testlet variants using the generator showed that the physics testlet templates as templates in AIG were functional. The characteristics of the testlet templates were a sub of the features of the generator. The physics testlet template used three questions from the 10 item templates provided by generator. From the generator feature checklist table, physics testlet templates could be entered manually or uploaded templates in Word documents (.docx), image files for templates were uploaded separately, templates used images in stem and options, templates used numerical expressions or formulas on the stem and options, the template used the item keys that were fixed and conditional, and the item keys were single and multiple. The Physics testlet templates used a single testlet question key, namely the use of multiple-choice format testlet compiler questions with one correct answer. This single key was accommodated by the generator besides the use of multiple keys, namely the use of testlet questions with multiple-choice format with more than one correct answer.

BBT testing through controlled experiments during program execution of the 25 physics testlet templates was carried out to ensure that the physics testlet templates followed the functionality of the testlet generator. This test reveald the "demonstration of proper behavior" [10] from the generation of testlet variants based on physics testlet templates. The results of this test could be interpreted as providing evidence about the quality of the physics testlet templates from the re-engineering of the national-exam physics questions in the context of automatic item generation.

4 Conclusion

The success of validating the testlet templates as input and the variants of the physics testlets generated by the generator as output shows that the physics testlet templates have characteristics that match the features of the generator for automatic item generation. The testlet templates of the results of re-engineering of the national-exam physi]cs questions conform to the functional specifications required by the testlet generator. The AIG perspective of the physics testlet templates has been verified by the generation of the testlet variant with the generator.

The suitability of the characteristics of physics templates with the generator features can be used as a reference for developing testlet templates for other physics subjects or other domains of science. The optimization of generator's features can be further explored in the development of testlet templates such as the use of multiple-choice multiple answer formats and the use of nested manipulative variables.

References

[1] Choi J, Kim H, Pak S. Evaluation of Automatic Item Generation Utilities in Formative Assessment Application for Korean High School Students. J Educ Issues. 2018 Mar 19;4(1):68.

[2] Gierl MJ, Lai H, Hogan JB, Matovinovic D. A Method for Generating Educational Test Items that are Aligned to the Common Core State Standards. 2015;16:18.

[3] Higgins D, Futagi Y, Deane P. Multilingual Generalization of the ModelCreator Software for Math Item Generation. ETS Res Rep Ser. 2005 Jun;2005(1):i–38.

[4] Higgins D. Item Distiller: Text Retrieval for Computer-Assisted Test Item Creation. :21.

[5] Gierl MJ, Zhou J, Alves C. Developing a Taxonomy of Item Model Types to Promote Assessment Engineering. :51.

[6] Gütl C, Lankmayr K, Weinhofer J, Höfler M. Enhanced Automatic Question Creator – EAQC: Concept, Development and Evaluation of an Automatic Test Item Creation Tool to Foster Modern e-Education. 2011;9(1):16.

[7] Ferreyra MF, Backhoff-Escudero E. Validez del Generador Automático de Ítems del Examen de Competencias Básicas (Excoba). RELIEVE - Rev Electrónica Investig Eval Educ [Internet]. 2016 Apr 14 [cited 2021 Jan 1];22(1). Available from: https://ojs.uv.es/index.php/RELIEVE/article/view/8048

[8] Kosh AE. A Process for Automatically Generating Algebra Items. 2019;20:18.

[9] Rambhau MA, Gupta S, Phadatare M, Jadhav I, Vaibhav BO. An Automatic Generator of Multiple Choice Question with Radom Answer Key. 2020;7(6):6.

[10] Oliinyk B, Oleksiuk V. Automation in software testing, can we automate anything we want? :11.

[11] Hamza ZA, Hammad M. Testing Approaches for Web and Mobile Applications: An Overview. :8.

[12] Stoen SM, McDaniel MA, Frey RF, Hynes KM, Cahill MJ. Force Concept Inventory: More than just conceptual understanding. Phys Rev Phys Educ Res. 2020 Jan 27;16(1):010105.

[13] Wagner M, Basevi H, Shetty R, Li W, Malinowski M, Fritz M, et al. Answering Visual What-If Questions: From Actions to Predicted Scene Descriptions. In: Leal-Taixé L, Roth S, editors. Computer Vision – ECCV 2018 Workshops [Internet]. Cham: Springer International Publishing; 2019 [cited 2021 Jan 1]. p. 521–37. (Lecture Notes in Computer Science; vol. 11129). Available from: http://link.springer.com/10.1007/978-3-030-11009-3_32

[14] Gierl MJ, Lai H. The Role of Item Models in Automatic Item Generation. Int J Test. 2012 Jul;12(3):273–98.