# Priscilla Evaluation Pilot Study: A Rasch Measurement Analysis

**Elspeth McKay, Keven Asquith, and Eugenia Smyrnova-Trybulska**

## 1 Introduction

In the digital society the e-learning and microlearning is becoming more and more popular. This innovation could become more effective both for students and teachers.

The term microlearning describes learning in small amounts ("micro"). The microlearning module is small in size, focused and is easily digestible. Most often, it is three to five minutes long with short chunks of information focused on a specific topic or task. Learners no longer need to sit in long and boring seminars, lectures or presentations. Now they can find time to study in their busy schedule. Features of microlearning are: • *Conciseness:* • *Focus:* • *Autonomy:.* • *Variety:* • *Interactivity:* • *Flexibility:* (Mikhailov, 2018) as well as *Time, Content; Curriculum; Form; Process; Mediality; Learning type* (Hug, 2005, p. 4).

A good example of a tool and environment supporting the work of both students and teachers is the Priscilla educational platform based on microlearning and microcources developed under the auspices of the FITPED project (www.fitped.eu).

E. McKay · K. Asquith
Cogniware.com.au, Melbourne, VIC, Australia
e-mail: e.mckay@cogniware.com.au

E. Smyrnova-Trybulska (✉)
Department of Humanistic Education, University of Silesia in Katowice, Katowice, Poland
e-mail: esmyrnova@us.edu.pl



E. Smyrnova-Trybulska et al. (eds.), *Microlearning*,
https://doi.org/10.1007/978-3-031-13359-6_7

## 2 Background

The research study authors were involved in teaching the programming courses for their IT students and using microlearning for increasing the effectiveness of their learning and teaching. This work has already been reported showing important results relating to "the final four datasets that were analysed to confirm the suitability of automated assessment of the microlearning units as predictors of at-risk students and students" outcomes in the introductory programming courses (Skalka & Drlik, 2020). In another research study, an international team of authors describe their research proposing a conceptual framework focused on the comprehensive training of future programmers using microlearning and automatic evaluation of source codes to achieve immediate feedback for students (Skalka et al., 2021).

The study, conducted by Draxler (2020) researched the environment-triggered microlearning and stressed that "in ubiquitous (micro)learning, any place is a potential learning environment, be it a couch at home or a noisy place in public" (Draxler, 2020, p. 1).

"One such innovation gaining traction is Microlearning, which offers learning opportunities through small bursts of training materials that learners can comprehend in a short time, according to their preferred schedule and location" (Gill et al., 2020, p. 780). The research "explored the potential of Microlearningwithin design education and how it can be implemented into the Product Design Manufacture programme at University of Nottingham Ningbo China to support teaching instruction and enhance the student learning experience post-COVID-19. (Gill et al., 2020, p. 780).

Other research devoted "the context-sensitive microlearning of foreign language vocabulary on a mobile device" (Beaudin et al., 2007, p. 55). "Phrases were presented on average 57 times an hour; this intense interaction was found to be acceptable even after extended use" (Beaudin et al., 2007, p. 55).

Authors Javorcik and Polasek (2019) in their research analysed the various "forms of eLearning: mobile learning, blended learning, adaptive eLearning or eLearning with gamification elements." They stressed that "the term microLearning had also becoming established" and also emphasized that "the greatest advantage of microLearning was that it is not dependent on the technology that the student was using" (p. 254). Researchers described the study "in order to be able to determine the potential of microLearning, pilot microLearning courses had been created, which were then compared with existing eLearning courses" (p. 254).

## 3 Research Design

This chapter represents the results of a pilot-study involving 26-programming student participants. The online questionnaire design used a five-level Likert-scale (1, 2, 3, 4, and 5) by Eugenia Smyrnova-Trybulska and conducted by Irena Polak in framework programming classes). According to the following criteria, the questionnaire's instructions were to rate the Priscilla environment values on a scale from 1 to 5, where one meant negative, and five meant positive. There were three groups of test-items identified as (1) Substantive value; (2) Didactic value; and (3) Technical value (see Appendix). These groups also identified 'required' participation.

An educational researcher's key focus is to design reliable test instruments for measuring performance outcomes in particular settings. In this sense, the reliability property (or behaviour) of the set of attitudinal aspects of the Priscilla programming environment, describe how consistent or error-free the measurements were (Mulyani et al., 2020; Frisbie, 1988). Such test-measurement results can show whether the agreement-levels achieved represent whether or not, these agreements occurred and whether the Priscilla environment attributes were useful for that particular set of participants. It is helpful to understand participant reactions to the questionnaire-items. For instance: consistently low agreement levels across the cohort may indicate the need for remedial instructional strategies (for particular participants). Or where questionnaire-items were not showing much agreement among others, requiring additional Priscilla strategy-solving activities suggestions to stretch those participants cognitively.

It is equally important to ascertain the testing instrument has (knowledge) construct validity to represent the expected Priscilla knowledge/trait experienced by the higher education community involved in this study. In this case, the researcher needs to understand the anticipated scope of the Priscilla knowledge and skill development trajectory in so far as checking the intent of the questionnaire test-items with established experts in the field, and that the test results on a particular test reflect the expected results on another relevant test. Validity can, however, mean different things, based upon the supporting evidence. Therefore, it is essential to understand the need to establish the Likert-level characteristics' interpretation of the questionnaire's rationale.

## 4 Results

The results of the questionnaire were analysed using the RUMM2030 Prof software application. This unidimensional psychometric Rasch measurement tool developed at the University of Western Australia (RUMM-2030, 2015), presents an opportunity for conjoint measurement of participant ability and latent trait (knowledge constructs) on the same linear scale divided into equal intervals along the scale (Andrich, 2011). An essential feature of a Rasch analysis is to ensure that the data fit the model. Other measurement models follow the item response theory (IRT), where the researchers search for a model to fit the data.

## 5 Data Analysis

The first step of the analysis was a manual (or visual) check of the data-file. The left-hand table below shows the 26-participants' IDs (marked as 1–26 in the left-hand column) and the remaining columns represent their Likert-score for each questionnaire/test-item across the row. The scored-outcomes for participant-6 revealed all fives (meaning complete agreement with all Priscilla questionnaire/test-items) and participant-17 mostly ones (indicating lack of agreement). These scored-outcomes stand out as being anomalous and require further review in the analysis. The remaining participants showed some variance in consensus across the questionnaire/test-items. The analysis needed to determine participants' overall agreement level to the Priscilla questionnaire/test-items, and whether individual test-item responses had expected outcomes, or had anomalous distributions. The right-hand table below displays some of the initial Rasch measurement analysis design, with 998-iterations resulting in 72 of the 105-parameters converging to a reliable outcome.

```
1  54444554554453443355353555b
2  45554544455432445443344455a
3  23343533453332335555154455a
4  22344555555454452245241555a
5  44555544554543554533354555a
6  55555555555555555555555555a
7  54444444454444443344433455b
8  55454534353553354551453455a
9  44555444454543454454433455a
10 44444544444444444533133455a
11 44454444422544441155222455b
12 42254544434444443555342355a
13 43442423344243444444443355b
14 55555555355553555553153555b
15 33344544523552444555253455a
16 55555555544554545545555555b
17 24334411111111111111111111a
18 22112512252222223331443125a
19 23335554554524443455451455a
20 55445545355453454553551555b
21 44434344213232343341143215b
22 53345544555452553345551255a
23 54444555455453444444554555b
24 55555555555553554555352555a
25 55455555555543443455443455a
26 55555555555555551153153555b
```

```
** PERSON  FACTORIAL  DESIGN
Person Factors            None entered

** ESTIMATION  DETAILS
Item Converge Limit       0.00010
Iterations [Items]        998 [only 72 parameters converged]
Iterations [Persons]      3
Lowest score              1
Highest score             130
Person Converge Limit     0.01000
Person Estimation         Weighted Maximum Likelihood method
Extreme Pers Criterion    0.220

** PARAMETERISATION
PC's requested            Location, Spread, Skewness and Kurtosis
Thresholds                131: Unequal across the 27 items
```

## 5.1 Person-Item Interaction

The proverb that a picture speaks a thousand words communicates the initial research results. The starting point is to examine the questionnaire/test results in terms of the individual participant/person Likert agreement-scores relative to each participant's agreement-score and each questionnaire-item's behaviour relative to each questionnaire-item on a common unidimensional scale, shown in the Person-Item Location Distribution below (Fig. 1). Summary statistics identify the unidimensional Rasch logit-scale location and fit residual statistics. There were 26-questionnaire participants. Overall, there were 27-items: 26-questionnaire items, each using a five-level Likert-scale, and one free textual questionnaire input response.

The threshold distribution for participants (red boxes above the Logit scale, Fig. 1) shows a skewed distribution with a mean value of 3.323 and a standard deviation of 1.348.

The questionnaire/test-item distribution is shown in blue boxes under the logit scale (Fig. 1) and indicates a much lower mean (0.000) relative to the person distribution on the same scale.

The summary statistics table from RUMM2030 indicates the participant distribution has a positive skewness of 1.118 (to the right of the mean) and a large Kurtosis of 5.477. Cronbach's Alpha Coefficient is 0.947 and the Rasch measurement Person Separation Index is 0.927, both indicating a good correlation to the model's expected values. The fit residual mean is only 0.074.

The questionnaire/test item distribution has a Std Dev of 2.542. The tail skewed to the left, with Skewness of −1.228 and a Kurtosis 1.489 indicating a broader distribution than the participants. The fit residual: mean is larger than for the participants at 0.423.

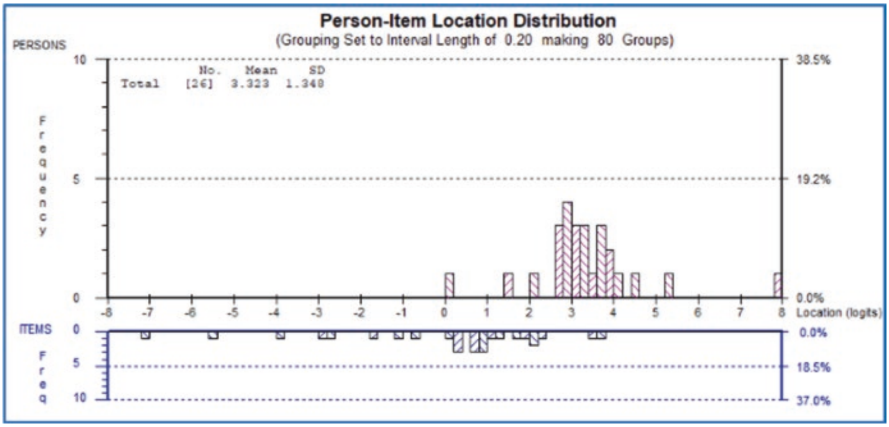
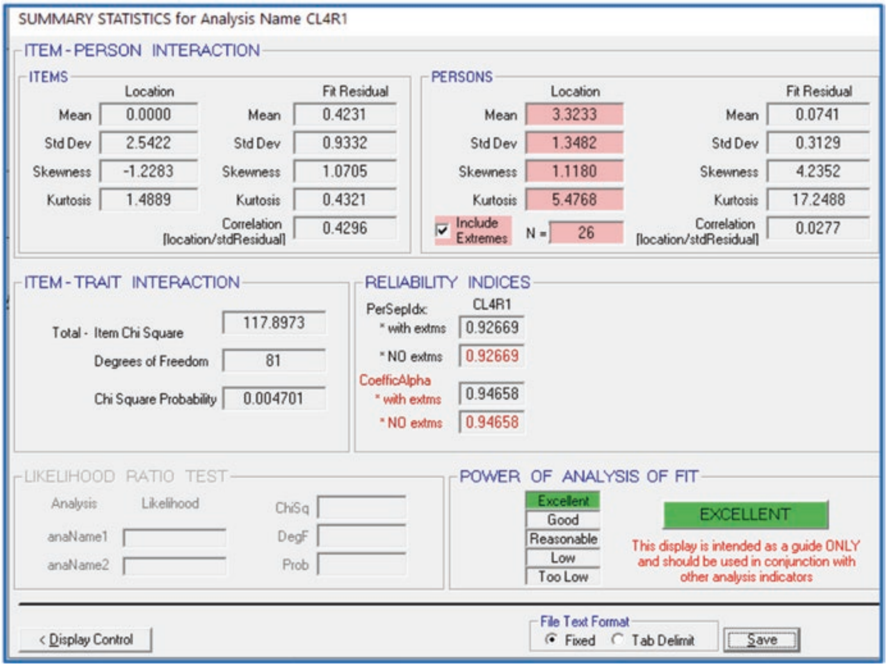
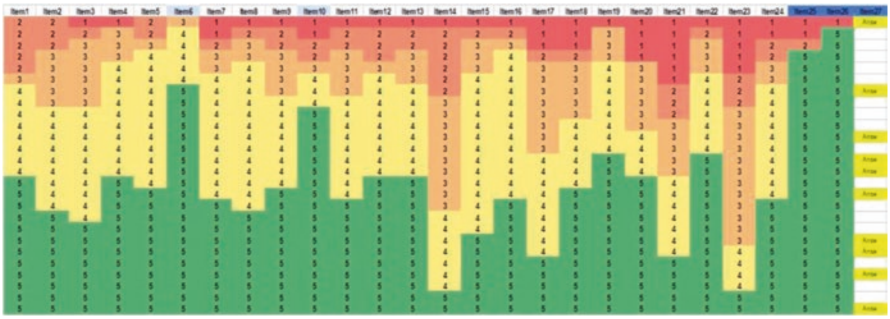



**Fig. 1** Person-item location distribution – Full data set




However, as identified in the data file's visual check, some participants and questionnaire/test-items may need further review. The data-file is redisplayed in the spreadsheet below to review questionnaire/test-items' ability to discriminate a range of responses. The response distribution for each test-item is colour-coded to identify the range of the 26-participants answers. Item-27 was a free text entry item, and therefore needs to be analysed separately to this Likert-scale analysis. Items-25 and 26, indicate that almost all participants answered the item at '5,' and consequently revealed they have a limited scored-value in discriminating response distribution. Test-items-6 and 10, also have a limited range of responses (again a large majority answered the test-item at '5') and needs further review.



The Rasch measurement item characteristic curve (ICC) was used to evaluate which of the above four test-items (6, 10, 25 and 26) should be removed from the analysis. The ICC identifies class intervals that show an item's relative ability to discriminate among adjoining knowledge constructs (traits) along a linear scale. Rather than compare individual participant scores against the expected model curve, the Rasch measurement model divides the sample into classes, in this case, quartiles. So it determines the mean logit score of each 25% of the participant sample (1.929, 2.900, 3.240 and 4.031). As you will see in the expected value plots for individual questions, these red marks on the person logit scale do not change. However, the group's observed value (the black dot within the plot) does vary for each question. Figure 2 shows the ICC for questions 25 and 26. The model's estimated distribution demonstrates that the expected value of five was reached before most of the class interval means. Therefore the item does not discriminate between classes and should be removed from the analysis.

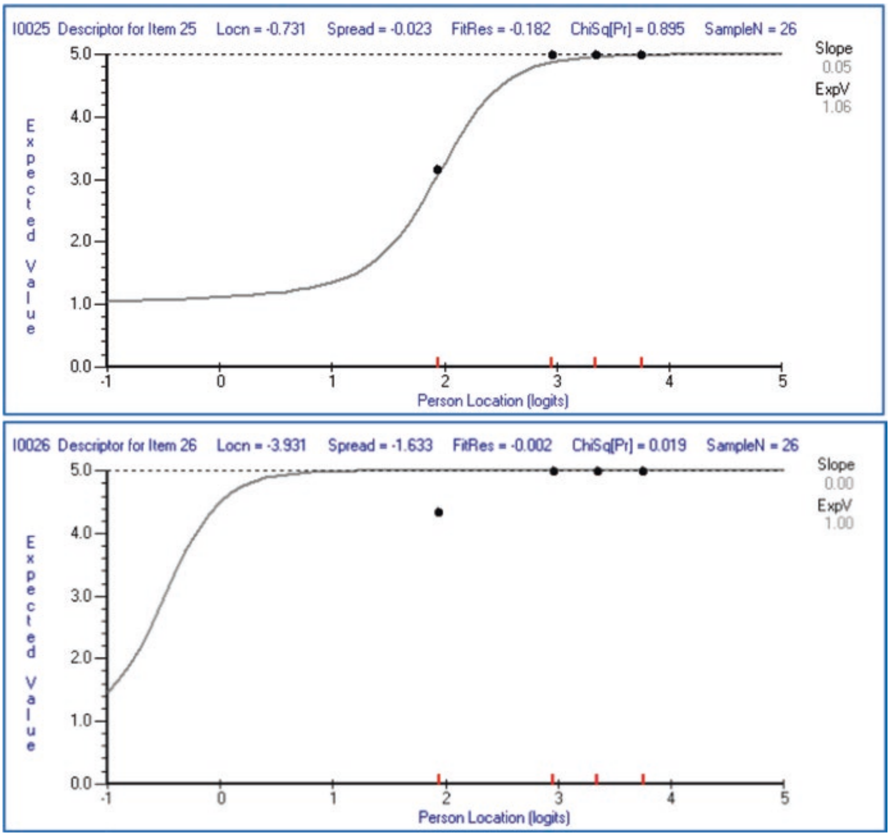



**Fig. 2** ICC for items 25 and 26

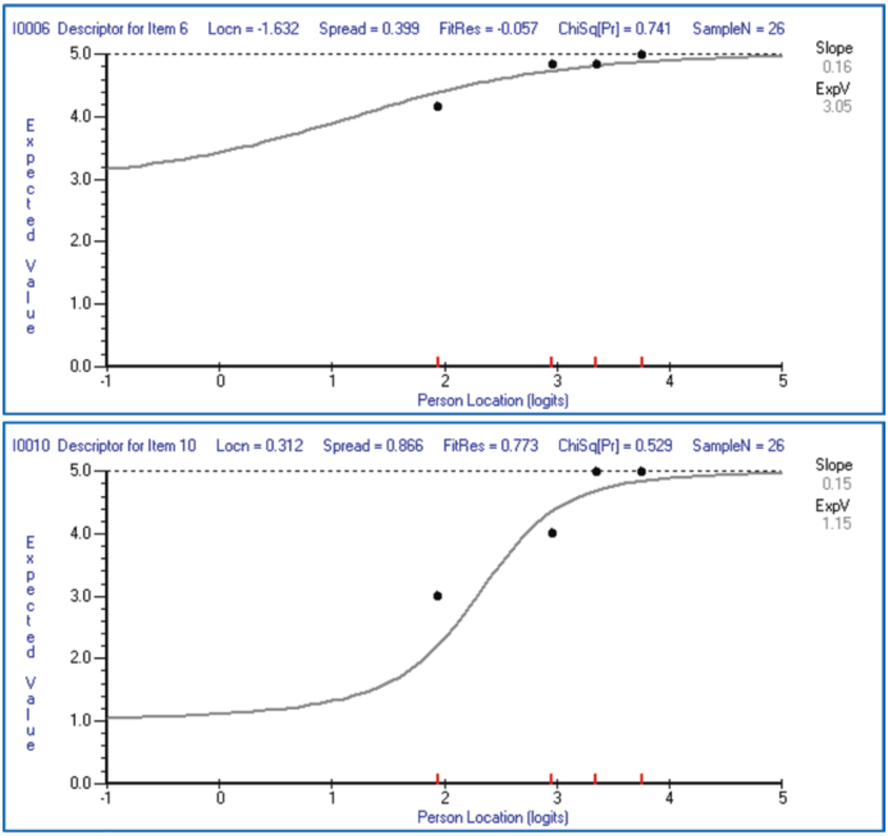



**Fig. 3** ICC for items 6 and 10

Conversely, the ICC for items-6 and 10 shows response discrimination and therefore should be retained in the analysis.

The RUMM2030 analysis was then repeated, with items 25, 26 and 27 removed. Figure 4 displays the Person-Item Location distribution with the three-items removed. The Person distribution mean has increased slightly to a value of 3.405 (from initial 3.323) and the standard deviation is 1.829 (previously 1.348). The Skewness to the right has risen to 3.009 (1.118) and Kurtosis increased to over 13 (previously 5.477).

The Fig. 4 distribution plot clearly illustrates the outlier participant with a score of over 11. Participant-6 identified as answering ‘5’ to every test-item. This extreme (participant) was skewing the distribution and was also removed from the analysis.

Figure 5 illustrates the Person-Item location distribution and summary statistics with the extremes removed from the analysis (now a sample of 25). Note that RUMM2030 did not remove participant-17 (answers mostly ‘1’) as their overall score did not fall outside the extreme criteria.

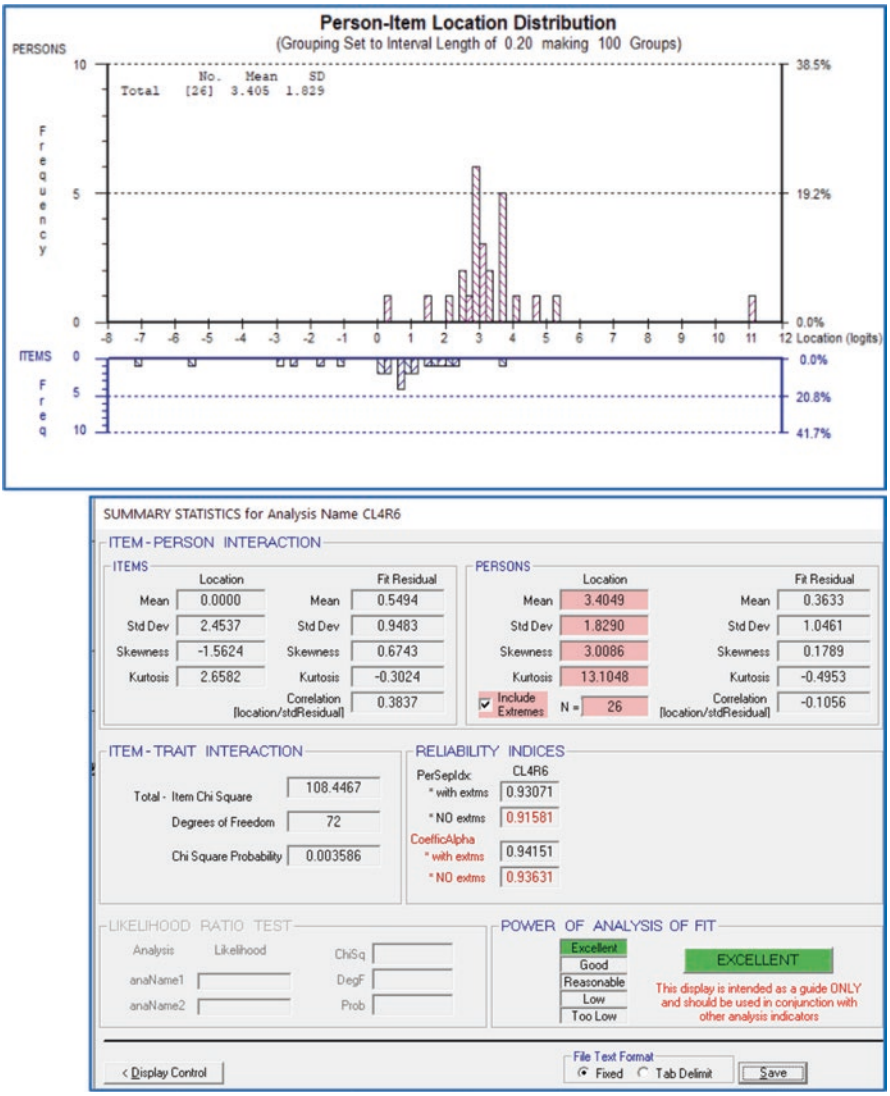



**Fig. 4** Person-item location distribution and summary statistics – With 3-items removed

The participant mean value has reduced to 3.097, almost 10% lower than the analysis with all participants (3.405) and the standard deviation has reduced to 0.967 (from 1.829). With the extreme removed, the participant results are nearing a normal distribution with a slight Skewness to the left of −0.533, and the Kurtosis has reduced to 2.498.

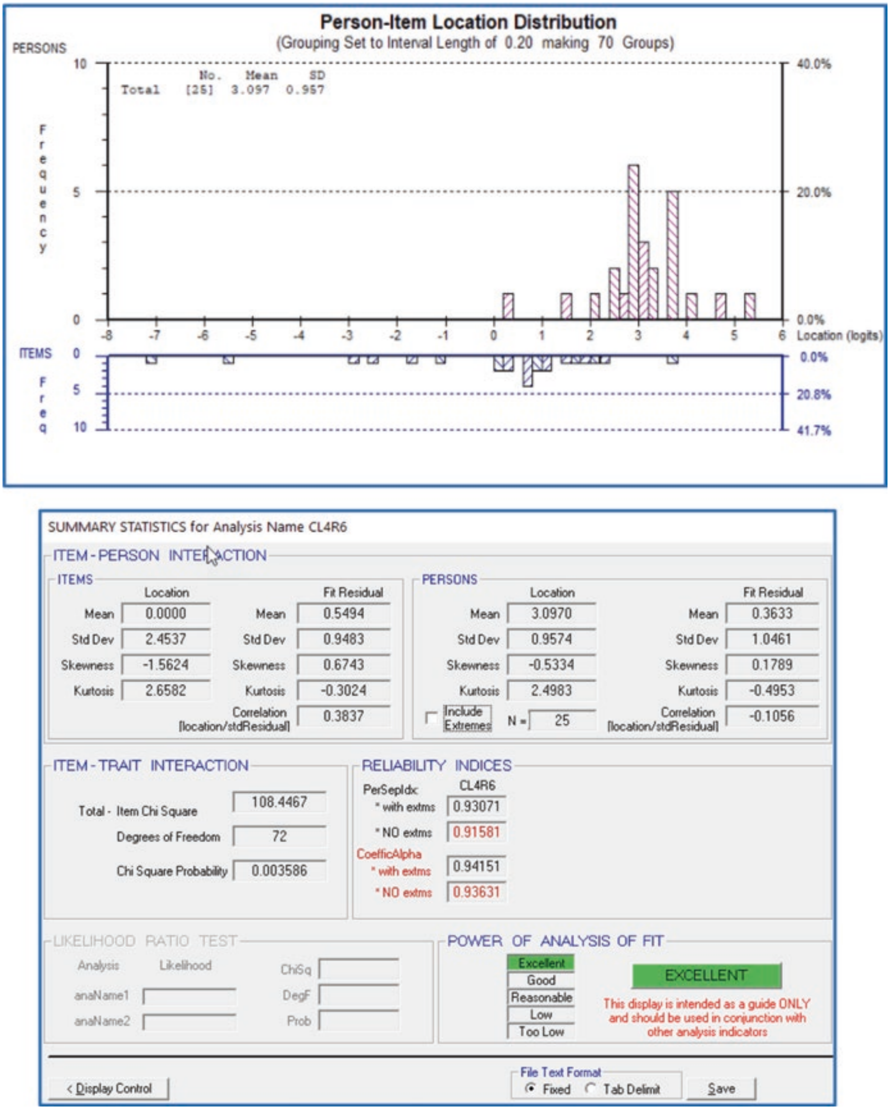



**Fig. 5** Person-item location distribution – Extreme participant removed

Cronbach's Alpha Coefficient is 0.936, and the Rasch model Person Separation Index is 0.916, both still indicating a good correlation to the model's expected values. The fit residual mean is only 0.363. Although this person-item threshold distribution map reveals a power of analysis fit as excellent, it should be noted that the display is intended as a guide only and should be used in conjunction with other analysis indicators (RUMM-2030, 2015).

The questionnaire item distribution, shown in blue histogram under the logit scale (shown earlier in Fig. 1), retains the mean of 0.000, while the Standard Deviation (SD) is 2.454 (the initial run-1 SD was 2.542). Similarly, there is an only minor variation in Skewness and Kurtosis between the initial analysis and the final analysis (with three questions and one participant removed).

Figure 6, an expanded version of the earlier person-item location distribution map, illustrates the different Likert-scaled agreement-levels for each question. The agreement levels generally progress upwards on the Logit scale from a 1-meaning a 'negative' attitude towards Priscilla and a 5-meaning a 'positive' response. Item-19 is highlighted to demonstrate the normal progression up the logit-scale. However, not all questions will follow this regular progression. It should be noted that the third level of the question often has a higher logit-value than level-4 of that same question, which may indicate participants selecting the midpoint of the Likert-scale if they were not sure of their response.

**Individual Item Analysis**

As previously mentioned, the questionnaire had 27-questions that grouped into three categories:

1. Substantive value (items 1–6)
2. Didactic value (items 7–16) and
3. Technical value (items 17–27)

Classical test theory (CTT) makes an assumed characterisation of a person through a total dichotomous (0 and 1) summed-score. Although the Rasch measurement theory (RMT) scores test-items in the same way as CTT, RMT uses the total person-score as the sufficient statistic due to the model (Andrich & Marais, 2019). There are marked differences between these two measurement techniques. CTT has no practical way to see whether test-items are working as expected. In contrast, RMT uses the probability that a *"person n with given proficiency βn responds correctly to an item i with difficulty σi "* (Andrich, 2010, p. 162).

The item characteristic curves (ICC) discussed previously, graphically display this fit concept by comparing the observed mean value of each RMT-class category

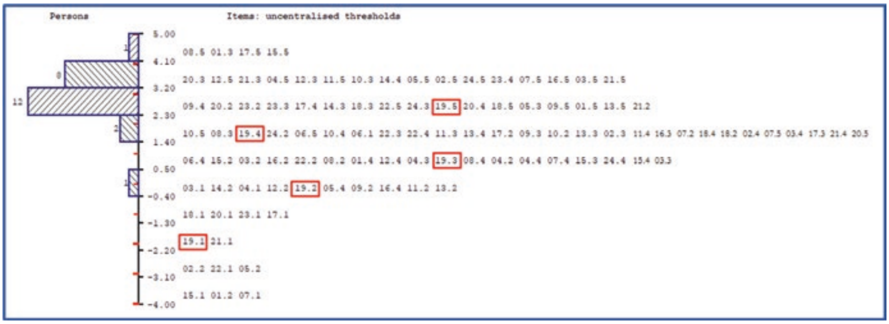



**Fig. 6** Final analysis item variable map

(black dots) against the model derived expected value curve for each question. The left-hand plot in Fig. 7 illustrates that test item-22 (asking if the environment is easy to run?) has an excellent fit to the model curve. The fit residual mean is only 0.293.

## 5.2 *The Category Characteristic Curve (CCC)*

The category characteristic curve (CCC) illustrates the probability of a participant responding to the category answer and the likely person's logit value. And so at any logit value, the sum of probabilities for each question will add up to 100%. Using Item-22 as a typical model distribution, showing a normal progression of Likert-scale responses, the right-hand of Fig. 7, illustrates that a participant with a logit value of a '1' would have a 40% probability of answering at the level- 2, a 30% probability of answering at level-1, a 20% probability of answering at level-3 and a 10% probability of answering at Likert level-4.

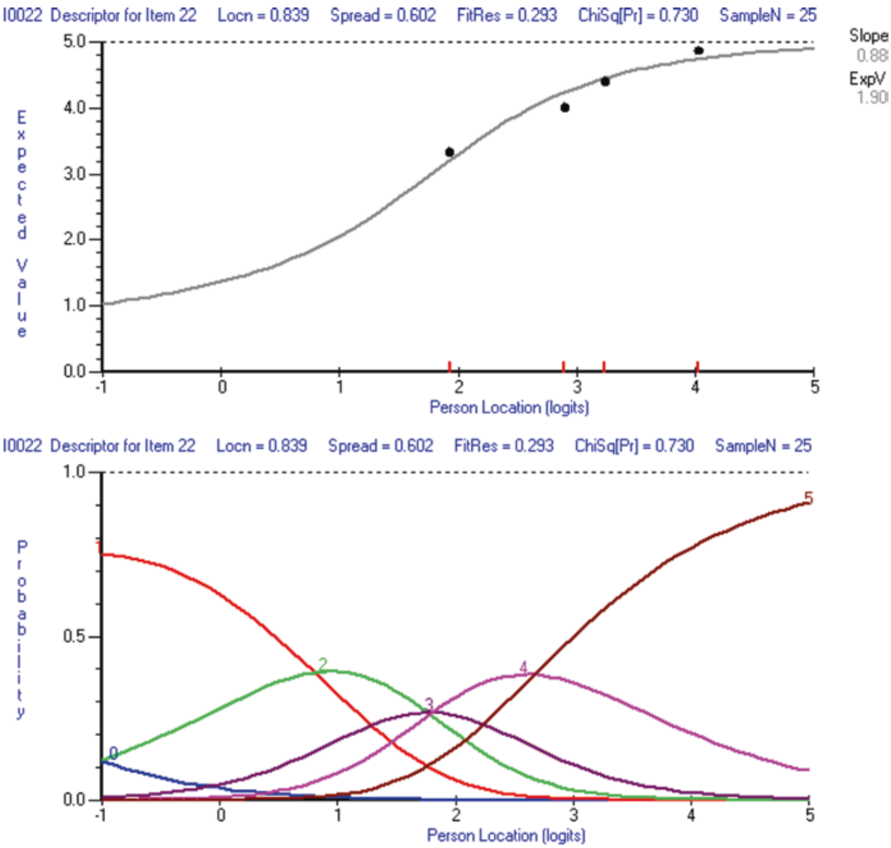



**Fig. 7** Final analysis item-22 ICC and CCC probability showing a good fit

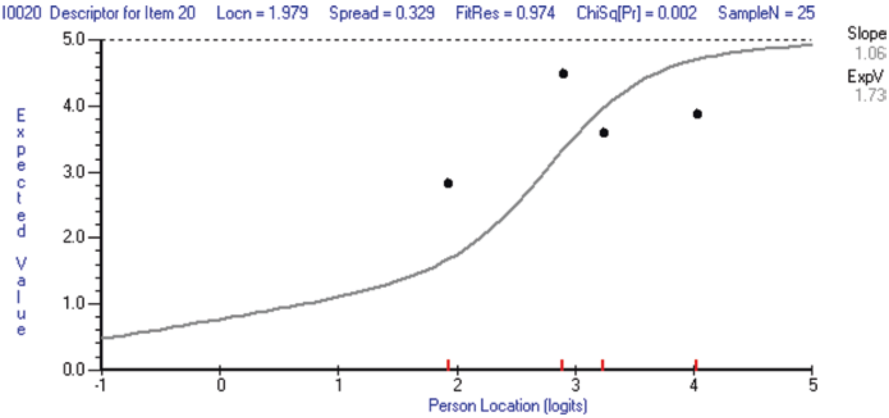



**Fig. 8** Final analysis item-20 ICC showing a poor fit

Looking at the agreement-scale endpoints, a participant with a logit value of −1.0 or lower has over a 75% probability of answering question-22 at the level-1 agreement-level (negative). A person with a logit value of '5' would have a 90% probability of answering question-22 at the level-5 agreement level (positive). The brown curve indicated that a person with a logit value of '1' had no chance of answering item-22 at the level-5 and that a logit-value of almost '3' is required to have a 50% probability of answering with a '5' response.

Conversely, Fig. 8 illustrated that test item-20 (asking whether sound has technical value) had a bad fit to the expected curve, with all observed class category, means plotted away from the model expected curve. The fit residual mean has increased to 0.974. This item may need further evaluation to determine if participants understood what was asked or whether the response indicated an anomalous distribution of attitude towards sound use.

In the final analysis, the remaining items have a model fit between these two endpoints and would be suitable for future experiments.

## 6 Conclusions

Following a Rasch measurement approach to data analysis has profound rewards, not possible through standard statistical evaluation. The interactivity of the person/item behaviours relative to each other and measured on a unidimensional (logit) scale, provides a robust tool to illustrate this complex environment. The category characteristic curve (CCC) and the item characteristic curves (ICC) are keys to understanding this powerful measurement model.

Rash measurement CCC plots each response's probability as a function of person proficiency and item characteristic curves (Andirch & Marais, 2019, p. 242). For instance: the category probability plot for item-22, which asked for an opinion on the Priscilla environment being easy to run, provides a robust comparison of students' probability to answer this question relative to their overall test-score. A participant with a '−1' overall logit-score has a 75% probability of having a negative agreement (a '1' on the Likert-scale). Whereas a participant with a logit-score of '3' has over 50% chance of having a positive agreement (a '5').

Conversely, item characteristic curves (ICC) reveal a fine-grained analysis of expected/observed score-values. For instance: test-item-20, which related to sound, did not fit the Rasch measurement model as observed mean-scores were erratic and did not follow the anticipated performance curve. Moreover, the fit residual mean indicated to the researcher to examine separately to determine whether students understood the question, or whether something else about using sound was responsible for the responses.

This pilot study data analysis resulted in the removal of two items with extreme scoring outcomes; item-25 – relating to hardware requirements for working with the environment not being excessive, had 23 of 26 participants score at level-5 (positive agreement), while item-26 – relating to cost not being excessive/free, had 25 level-5 answers (with only participant-17, who scored most items as a '1,' scoring differently). Consequently, removing these items afforded a more accurate and practical estimate. Examining the person-item statistics identified two outlier participants, who may have carelessly answered the questionnaire. Only one participant was removed from the final analysis. Evidence for this likely haphazard approach to participant scoring outcomes is seen by examining the Final Analysis Item Variable Map (often referred to as a Wright Map (Boone et al., 2014)); where not all items followed a regular progression along the five-level agreement Likert-scale.

There can be no doubt that running a pilot study has merits for validating the measurement instrumentation's reliability and the schedule of events before running the main experiment. In this case, the observation concerning the participant criteria/information on the five-category Likert-scale suggests detail regarding each of the five-level Likert-scale is necessary to elicit reliable outcomes from all test-items. To increase discrimination between participant's agreement levels, a seven-level Likert-scale could also be considered.

The one of the main aims of the FITPED Project was reducing the number of students because of failure to learn programming in the first years' study on IT specialisation via introduction Priscilla platform and elaborating above 2000 microcources of a programming language. Computer programmers are responsible for the following tasks: Translating program designs into code; Mastering computer languages; Borrowing from code libraries; Testing and troubleshoot programs; Using integrated development environments (IDEs) (Walter, 2020). The first research results showed that the students were generally good at evaluating the Priscilla environment. Simultaneously they should be additionally improved in several other issues, like: minor technical problems; correcting bugs and adding more theories in some sections before formulating queries (note: in some microcourses); sometimes the generator compiles the same tasks/exercises for execution. The authors of the Priscilla platform and the international team of other authors offer more than

2000 microcources that continue improving the environment and didactic tools using permanent monitoring and feedback.

**Acknowledgments** This research was funded by European Commission under the ERASMUS+ Programme 2018, KA2, grant number: 2018-1-SK01-KA203-046382 “Work-Based Learning in Future IT Professionals Education”.

## Appendix

**Fitped Survey** Evaluating the Priscilla Environment (students – Databases – UŚ) Evaluation of the Priscilla environment according to the proposed criteria.

Please rate the Priscilla environment values on a scale from 1 to 5 according to the following criteria – where 1 means negative and 5 means positive. * Required

1. **Substantive value***
    1. Substantive correctness
    2. Formulating information contained in the environment/microcourses
    3. Instructions
    4. Adequacy of the environment/micro-courses to the age of the learners
    5. The information contained in the environment/micro-courses takes into account the current knowledge
    6. Adherence to ethical standards, tolerance and gender policy in the program
2. **Didactic value***
    7. Individualisation of teaching
    8. The approach to the presented problems
    9. Ability to work with the environment/micro-courses of more than one person
    10. Is it possible to make multiple attempts to solve the problem in the program?
    11. The interactivity of the environment/micro-courses
    12. The environment/micro-courses are adjusted to the level of development of students
    13. The environment/micro-courses leave the lecturer free in terms of methodology
    14. The use of the environment/micro-courses causes that other teaching aids are not so effective
    15. Adequacy of information describing the environment to the package contents
    16. The environment/micro-courses contain content compatible with the text-book/script/teaching materials
3. **Technical value***
    17. Is the workflow in the environment efficient and clear?
    18. Is the environment “user friendly”?

19. Graphics
20. Sound
21. Does the environment/micro-courses predict user errors or mistakes?
22. The environment is easy to run
23. It is possible to change the way information is presented
24. The environment/micro-courses make the most (properly) use of the student's time
25. The hardware requirements for working with the environment are not excessive
26. Program cost (not (excessive), free)
27. Is the workflow in the environment efficient and clear?

4. **Please send your suggestions to improve and enhance the Priscilla environment and programming language learning micro-courses**

## References

Andrich, D. (2010). Rasch models. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 111–122). Elsevier.

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review Pharmacoeconomics Outcomes Research, 11*(5), 571–585.

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Springer Texts in Education.

Beaudin, J. S., Intille, S. S., Tapia, E. M., Rockinson, R., & Morris, M. E. (2007). *Context-sensitive microlearning of foreign language vocabulary on a mobile device*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Volume 4794 LNCS, Springer Verlag, 2007, Pages 55–72 European Conference on Ambient Intelligence, AmI 2007; Darmstadt; Germany; 7 November 2007 through 10 November 2007; Code 70979.

Boone, W. J., et al. (2014). *Rasch analysis in the human sciences*. Springer.

Draxler, F. (2020, October). Environment-triggered microlearning. *MobileHCI'20: 22nd international conference on human-computer interaction with mobile devices and services*. Article No.: 40, pp. 1–3 https://doi.org/10.1145/3406324.3409259.

Javorcik, T., & Polasek, R. (2019). Practical application of microlearning in education of future teachers. *Proceedings of the European conference on e-Learning*, ECELVolume 2019-November, 2019, pp. 254–259. *18th European Conference on e-Learning, ECEL 2019*; Aalborg University Copenhagen; Denmark; 7 November 2019 through 8 November 2019; Code 155850.

Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Instructional Topics in Educational Measurement (ITEMS)*. Spring, pp. 25–35.

Gill, A. S., Irwin, D. S, Ng, R. Y.- K., Towey, D., Wang, T., & Zhang, Y. (2020, December). The future of teaching post-COVID-19: Microlearning in product design education. *Proceedings of 2020 IEEE international conference on teaching, assessment, and learning for engineering*, TALE 20208, Article number 9368322, pp. 780–785. *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020*; Virtual, Takamatsu; Japan; 8 December 2020 through 11 December 2020; Category number CFP20TAL-ART; Code 167764.

Hug, T. (2005, May 6–8). *Micro learning and narration. Exploring possibilities of utilization of narrations and storytelling for the designing of "micro units" and didactical microlearning* *arrangements*. Paper presented at the fourth Media in Transition conference, MIT. https://www.researchgate.net/publication/237558117_Micro_Learning_and_Narration_Exploring_possibilities_of_utilization_of_narrations_and_storytelling_for_the_designing_of_micro_units_and_didactical_micro-learning_arrangements. Accessed 10 Sep 2017.

Mikhailov, V. (2018). Microlearning – Features, advantages and disadvantages. *Education*. https://edu-russian.ru/mikroobuchenie-osobennosti-preimushchestva-i-nedostatki.html

Mulyani, H., et al. (2020). Quality analysis of teacher-made tests in financial accounting subject at vocational high schools. *Jurnal Pendidikan Vokasi, 10*(1).

RUMM-2030 (2015). *RUMM2030 getting started manual*. WA, RUMM_Laboratory_Pty_Ltd.

Skalka, J., & Drlik, M. (2020). Automated assessment and microlearning units as predictors of at-risk students and students' outcomes in the introductory programming courses. *Applied Sciences, 10*(4566), 1–24. https://doi.org/10.3390/app10134566

Skalka, J., Drlik, D., Benko, L., Kapusta, J., Rodríguez del Pino, J.C., Smyrnova-Trybulska, E., Stolinska, A., Svec, P., & Turcinek, P. (2021). Conceptual Framework for Programming Skills Development Based on Microlearning and Automated Source Code Evaluation in Virtual Learning Environment. *Sustainability, 13(6),* https://doi.org/10.3390/su13063293 https://www.mdpi.com/2071-1050/13/6/3293 ISSN:2071–1050

Walter, J. (2020). *Learning to code? Strong language skills matter more than being good at math*. https://www.discovermagazine.com/mind/learning-to-code-strong-language-skills-matter-more-than-being-good-at-math