Research Article

A New Approach to Analyzing the Development of Domain-Specific Knowledge Among Undergraduate Medical Students Using Learning Scores

Jochen Roeper[1]*, Olga Zlatkin-Troitschanskaia[2]*, Verena Klose[1], Marie-Theres Nagel[2] and Jasmin Schlax[2]

[1]Department of Neurophysiology, Goethe University Frankfurt, Germany

[2]Department of Business and Economics Education, Johannes Gutenberg University Mainz, Germany

*Corresponding author: Olga Zlatkin-Troitschanskaia, Department of Business and Economics Education, Johannes Gutenberg University Mainz, Germany.

Abstract

Medicine is one of the domains in higher education with the highest demands [1]. The importance of acquiring domain-specific knowledge to understand the scientific rationales of practical work in medicine is considered vital [2], about the specific responsibilities associated with medical professions. Teachers in medicine are faced with specific instructional challenges, for instance, being required to teach and examine several hundred students with heterogeneous educational backgrounds and study preconditions [3]. Consequently, in medical education practice, Multiple-Choice (MC) tests are often used in examinations because of their high level of efficiency and practicability, and despite their well-known disadvantages such as construct-irrelevant bias and low levels of explanatory power [4-5]. In particular, the extent to which teaching contributes to the development of domain-specific knowledge cannot be measured by simple post-testing with MC tests and analyzing test sum scores only [6-7].

Thus, practical implications for teaching and learning in medicine can only be derived to a very limited extent. In this paper, we present a new approach to analyzing the development of medical knowledge and suggest that a design with pre- and post-measurements and the consideration of decomposed pre- and posttest scores, which we define as learning scores, can provide substantial additional information on medical students' learning over the course of their studies. This practicable approach can help to inform educational practitioners about students' difficulties in learning and understanding certain medical contents, as well as to uncover possible student misconceptions about medical concepts and models.

Keywords: Development of knowledge, Learning scores, Partial scoring, Knowledge in physiology, Medicine education, Practitioners, Psychology, Intelligence, Sociodemographic, Heuristic

Introduction

Usually, the development of domain-specific knowledge is measured by comparing two measurement points or by using a difference score. However, this approach does not sufficiently reflect the learning process and it is difficult to derive implications for teaching practice. In some studies, an alternative approach has been presented, in which knowledge development can be divided into different types of learning: positive (knowledge gain), negative (knowledge loss), retained (stable knowledge) and zero learning (no learning) [8-9]. However, this approach is also limited to domain-specific tasks where distractors cannot be classified as false or correct with absolute certainty. This is often the case in tasks representing specific medical content.

Learning and developmental psychology research suggests that learning processes occur in a non-linear fashion by crossing thresholds [10-12], which also includes the acquisition and rejection of misconceptions [13]. Thus, tasks for validly assessing acquired domain-specific knowledge should be more appropriate to the developmental character of learning processes and move away from simple binary correct-incorrect coding and rather use more differentiated scorings with sub-points [14-16]. Such differentiated

learning scores enable the consideration of different learning patterns that go beyond the previously mentioned established approaches.

A New Approach to Learning Scores

The approach presented here offers an innovative complementary method for analyzing pretest and posttest responses from an MC test to assess medical students' learning processes, in this case based on an example from physiology. Using an innovative score decomposition approach, the seven identified student response patterns (Table 1) provide indicators of four student learning outcomes that, for the purposes of this study, are described as completed learning, progressive learning, incomplete learning and interrupted learning. The decomposed learning scores can be used to analyze how they affect the MC test scores and for investigating factors at the item and student level that are likely to influence students' understanding of concepts in physiology.

Table 1: Taxonomy of learning.

	Name of Category	Defined as
	Completed learning	3
(partial) learning progression	Progressive learning	2
	Incomplete learning	1
	Stagnation/Standstill	0
(partial) learning regression	Beginning learning regression	-1
	Progressive learning regression	-2
	Knowledge loss/Interrupted learning	-3

In our study, such a pre-post-design was implemented to consider the development of the students' domain-specific knowledge before and after attending physiology lectures. A short domain-specific knowledge test (12 items) in an MC format with five response options each was developed and, as described below, scored (with subpoints) by experts. This allows for a differentiation of the following types in students' knowledge development for each test item (Figure 1). The learning patterns can be divided into categories as indicated in Table 1.
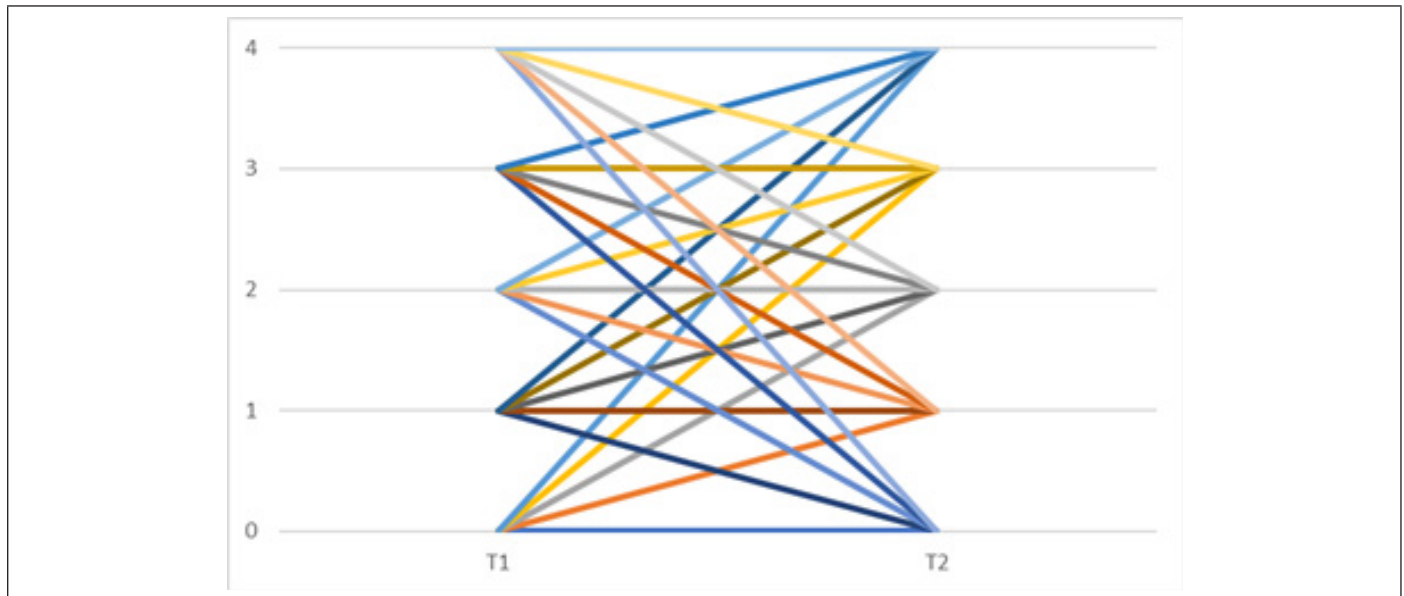

Figure 1: Possible learning patterns for a student on a test item (4 points-full correct solution).

In our study, we focus on the following two research questions (RQ):

I. RQ1: How can the learning process be described regarding the development of domain-specific knowledge between both measurement points-before (t1) and after (t2) attending the physiology lectures when using our decomposed approach?

II. RQ2: Which learning patterns become evident at the individual and item level?

In addition, based on our investigation in the study reported here, we particularly aim to explore which implications for medical educational practice can be drawn from the student learning scores in terms of improving teaching and learning.

Study Design, Sample and Test Instruments

Within the framework of a pre-post study, surveys were conducted at the beginning (t1) and end (t2) of the summer semester 2019, six months apart. Second semester medical students attending physiology lectures at one German university were surveyed (N in t1=134; N in t2=132; matched sample, i.e. students who participated in both measurements, N=97. The surveys took place at the beginning of the first and penultimate lecture of the semester and lasted about 25 minutes. In addition to the participants' domain-specific knowledge in physiology, intelligence, sociodemographic data such as gender, language and previous education as well as use of print and digital learning materials and sources (such as text books, scripts, learning platforms and study groups) over the course of their medical studies were collected. As an incentive for participation, the students were given the opportunity to view their test results online after both measurements and to obtain individual feedback on their knowledge development.

At the first measurement point, 134 medical students took part, most of whom were in their second semester (m=2.03, SD=0.10). More than two thirds (N=97, 72.39%) of the participants were female and most of the students (N=116, 86.57%) stated that their mother tongue was German. Only one fifth (N=28, 20.9%) of the participants had not completed an advanced science course at school. At the second measurement point, the same students were surveyed again. At both measurement points, domain-specific knowledge was assessed using a newly developed test containing twelve text-based questions about physiology in an MC format, each with five possible responses. The subsequent evaluation of the student responses was based on a partial credit system, as described above, meaning that the participants were given between 0 and 4 points per task depending on the chosen distractor, which had been ranked by experts beforehand. Each task was rated by at least two experts and the scoring results show a high interrater agreement.

The participants' intelligence was assessed at the first measurement point using 20 items from the figural-spatial scale "Figure Selection" from a German intelligence test [17]. The participants had seven minutes to determine which of five given figures can be formed by putting together different figure pieces. To measure the students' use of learning sources, two scales were used at both measurement points to determine the frequency of use of 11 media sources for learning adapted for this study, for an original version, Maurer et al. To evaluate how frequently students use these sources, the participants were asked to respond on a scale from 1 (never) to 6 (multiple times a day) to the two questions "In your last semester at university, how many times a week did you use different types of media to find information about physiology topics/to prepare yourself for lectures and exams in physiology?".

They had the option of adding additional sources if applicable.

As control variables, the scale on information overload the scale on need for cognition as well as two further scales were used at t2 to measure the processing of information over the course of studies. Heuristic and systematic information processing were measured with three items each, using a 6-point scale from 0 ("does not apply at all") to 5 ("does fully apply"), validated by Schemer et al. and following the Risk Information Seeking and Processing model [18,19].

Statistical Procedure

After testing the preconditions of one-dimensionality of the domain-specific test in physiology used in this study (by using confirmatory factor analysis) for the two types, learning progression and learning regression, the analysis of the knowledge sum score of t1 and t2 as well as the testing of a mean value difference via a t-test provided first insights into the development of knowledge from t1 to t2. To investigate underlying learning processes, learning patterns were generated (as described above) and then evaluated at the individual and item level.

The development from t1 to t2 was first generated item-specifically. For the individual-centered analysis, mean scores were then generated for the different learning patterns. Multiple regressions were performed to analyze possible predictors of the learning scores, whereby measured cognitive traits (such as intelligence) were integrated in addition to socio-demographic variables (such as gender). The analyses of the research questions were carried out with Stata Version 15 [20], while MPlus Version 7 [21-23] was used to assess the students' preconditions and [24] latent characteristics.

Summary of Preliminary Results

I.   RQ1: How can the learning process be described regarding the development of domain-specific knowledge between both measurement points-before (t1) and after (t2) attending the physiology lectures when using our decomposed approach?

Thanks to the expert rating of the domain-specific test tasks, it became evident that for about half of the tasks, for instance, tasks where knowledge about physiological facts is required, a clear response should be expected, so that the experts recommended adhering to the binary rating-0 (wrong response), 1 (right response). For the remaining test tasks, for instance, with numbers or ranges of values in the distractors, where partial scoring was theoretically possible, the scoring was performed based on the empirical data. In these tasks, the expected learning patterns for the two types, learning progression and learning regression, could be differentiated in this sample [25,26].

II.   RQ2: Which learning patterns become evident at the individual and item level?

Significant differences in learning scores could be found at both item and student level. At the item level, for instance, there was one item in which most students evidently had difficulties understanding the underlying physiological concept and where learning scores showed a high value of "interrupted leaning". At the individual level, differences in learning scores about gender were found, indicating more learning patterns in the area of learning progression among female students compared to male students [27,28].

Conclusion

The results regarding the two research questions presented indicate both advantages and limitations of our newly developed decomposition approach to analyze student learning. The advantages can be seen in the high practicability of this approach, which allows for differentiated analyses of the development of knowledge over the course of studies at both individual and item level. Analyzing learning scores enables us to uncover difficulties and misconceptions in students' understanding of medical concepts and models as well as in their application in practical contexts, for instance, solving a numerical task, so that as a result, medical education practitioners can effectively tailor their mode of instruction. In this sense, this new approach can significantly enrich formative assessment in medical classrooms and can be used both as a tool for medical educators to improve their instruction and as a tool for students to optimize their learning strategies.

The limitations of the approach lie in its applicability to only specific medical content, i.e. tasks where a partial scoring is theoretically meaningful. When developing a domain-specific test in medicine, an expert rating would have to be carried out to enable objective partial scoring and to examine the interrater agreement to ensure reliable measurement. In addition, further longitudinal analyses with more than two measurement points are strongly recommended to carry out a more comprehensive validation following the internationally established Standards for Educational and Psychological Testing [24]. The longitudinal data would also allow for the application of more sophisticated statistical analyses such as using cognitive diagnostic modelling. Based on the preliminary results, further work on the theoretical-conceptual analyses such as definition and description of thresholds for the different medical concepts is also required in future studies.

References

1. Fischer MR, Bauer D, Mohn K (2015) Finally finished! National competence-based catalogues of learning objectives for undergraduate Medical Education (NKLM) and dental education (NKLZ) ready for trial. GMS Time education for medical education 32(3): 35.

2. Wijnen M, Der Schaaf VM, Nillesen K, Harendza SO, Cate T (2013) Essential facets of competence that enable trust in medical graduates: a ranking study among physician educators in two countries. Perspect Med Edu 2(5-6): 290-297.

3. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR (2010) The role of assessment in competency-based medical education. Med Teach 32(8): 676-682.

4. Lau P NK, Lau SH, Hong KS, Usop H (2011) Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests. Jou Edu Techno Society 14(4): 99-110.

5. Smith B, Wagner J (2018) Adjusting for guessing and applying a statistical test to the disaggregation of value-added learning scores. Journal of Economic Education 49(4): 307-323.

6. Emerson T, English LK (2016) Classroom experiments: Teaching specific topics or promoting the economic way of thinking. Journal of Economic Education 47(4): 288-299.

7. Méndez Carbajo D, Wolla SA (2019) Segmenting Education Content: Long-form vs. Short-form Online Learning Modules. American Journal of Distance Education 33(2): 108-119.

8. Walstad WB, Wagner J (2016) The disaggregation of value-added test score to assess learning outcomes in economics courses. Journal of Economic Education 47(2): 121-131.

9. Happ R, Zlatkin TO, Schmidt S (2016) An analysis of economic learning among undergraduates in introductory economics courses in Germany. The Journal of Economic Education 47(4): 300-310.

10. Ravand H, Robitzsch A (2018) Cognitive diagnostic model of best choice: a study of reading comprehension, Educational Psychology 38(10): 1255-1277.

11. Alexander PA, Jetton TL, Kulikowich JM (1995) Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. Journal of Educational Psychology 87(4): 559-575.

12. Meyer JHF, Land R (2003) Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising. Rust C (ed.). Impr Student Learning-Theory Prac pp.412-424.

13. Villafane SM, Loertscher J, Minderhout V, Lewis JE (2011) Uncovering student's incorrect ideas about foundational concepts for biochemistry. Chemistry Education Research and Practice 12(2): 210-218.

14. Frary RB (1989) Partial-Credit Scoring Methods for Multiple-Choice Tests. Applied Measurement in Education 2(1): 79-96.

15. Bauer D, Holzer M, Kopp V, Fischer MR (2011) Pick-N multiple choice-exams: a com-parison of scoring algorithms. Advances in health sciences education. Adv Health Sci Educ Theory Pract 16 (2): 211-221.

16. Simon AB, Budescu DV, Nevo B (1997) A Comparative Study of Measures of Partial Knowledge in Multiple-Choice Tests. Applied Psychological Measurement 21(1): 65-88.

17. Liepmann D, Beauducel A, Brocke B, Amthauer R (2007) IST 2000 R Göttingen: Hogrefe.

18. Griffin RJ, Neuwirth K, Giese J, Dunwoody S (2002) Linking the heuristic-systematic model and depth of processing. Communication Research 29(6): 705-732.

19. Eagly AH, Chaiken S (1993) The psychology of attitudes. Harcourt brace Jovanovich college publishers, USA.

20. StataCorp (2017) Stata Statistical Software: Release 15 Computer software. College Station TX: StataCorp LLC.

21. Muthén LK, Muthén BO (2015) MPLUS (Version 7) Computer software. Muthén, Muthén (eds.), Los Angeles, USA.

22. Bauer D, Holzer M, Kopp V, Fischer MR (2011) Pick-N multiple choice-exams: a com-parison of scoring algorithms. Advances in health sciences education. Adv Health Sci Educ Theory Pract 16(2): 211-221.

23. Shakhar GB, Sinai Y (1991) Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies. Journal of Educational Measurement 28(1): 23-35.

24. Standards for educational and psychological testing. 2014th (edn). American Educational Research Association (AERA). Washington DC, USA.

25. Owen AL (2012) Student Characteristics, Behaviour, and Performance in Economics Classes. International Handbook on Teaching and Learning Economics. Hoyt GM, McGoldrick KM (eds.), Northampton, Edward Elgar Publishing, USA. pp.341-350.

26. Parker K (2006) The effect of student's characteristics on achievement in introductory microeconomics in South Africa. South African Jou of Eco 74(1): 137-149.

27. Pohl S, Gräfe L, Rose N (2014) Dealing with Omitted and Not-Reached Items in Competence Tests: Evaluating Approaches Accounting for Missing Responses in Item Response Theory Models. Educational and Psychological Measurement 74(3): 423-452.

28. Silva L O, Zárate L E (2014) A Brief Review of the Main Approaches for Treatment of Missing Data. Intelligent Data Analysis 18(6): 1177-1198.