

Research Note

Examining the Characteristics of Test Items in State Wide Exit Examinations: An International Comparison

by Mirko Krüger (mirko.krueger@uni-due.de)

Abstract: Most Organisation for Economic Co-operation and Development (OECD) states have installed state wide exit examinations at the end of upper secondary education (ISCED 3A). These examinations are assumed to set standards for learning outputs, and thus secure comparable results and educational quality. However, little is known about the item characteristics in the context of state wide examinations. Against this background, this PhD study investigates biology items in state wide exit examinations at the end of general upper secondary schooling in six European OECD states. Focusing on structure- and content-related dimensions by using a concept-oriented rating manual, the study shall disclose the general examination practice as well as trends concerning structure and content of the items and whether they possess more or less innovative characteristics regarding the concept of scientific literacy. By identifying good practice and innovative trends across the states, the results of the study can also contribute practically to the development of biology tasks designed to improve biology instruction.

Introduction

Most European Organisation for Economic Co-operation and Development (OECD) states have installed state wide exit examinations at the end of upper secondary education (ISCED 3A). It is assumed that they create a more standardized framework for student graduation from general upper secondary schooling than do school based exit examinations (Klein, Kühn, van Ackeren and Block, 2009). Having students pass the same exit examination is supposed to secure the comparability of student achievement within an administrative area, as every student has to show certain competencies based on a common curriculum for each area to graduate from upper secondary education and to gain access to universities. In addition to this “traditional” function of exit examinations, they are nowadays also increasingly used to positively affect learning and teaching processes and outputs, and to make sure that new syllabi and innovative instructional methods are implemented quickly and comprehensively (Bishop, 1998; Maag Merki, 2010).

However, little is known about how exit examinations actually affect schooling and the existing findings are very inconsistent (Baumert and Watermann, 2000; Vogler and Carnes, 2009; Maag Merki, 2010). Beyond that, a systematic international comparison of state wide exit examinations in several OECD states reveals different standardization levels of organizational conditions. Furthermore, the supposedly uniform label “state wide exit examinations” covers examinations of rather heterogeneous designs (Klein *et al.*, 2009). Therefore, it must be questioned whether all of these examinations are able to fulfill the expectations linked to state wide exit examinations, or whether different designs also cause different effects.

Item characteristics in the context of state wide exit examinations

Test items play an important role in the context of state wide exit examinations. The use of tests with uniform items serves as a comparable assessment and classification tool of students’ performances at the end of general upper secondary education. In general, the tests show the knowledge and competencies students have gained at the end of their studies

regarding the national standards. In addition, tests can be used to check whether students have the necessary skills for the academic or vocational tracks they will go on to after graduation. From the perspective of *Educational Governance* (Altrichter, Brüsemeister and Wissinger, 2007), items in state wide exit examinations are also an essential element with which the instruction in schools can be influenced, and the implementation of innovative contents and methods can be prompted, as the tests force teachers to cover certain content and instructional methods.

Regarding the high hopes connected to the introduction of these examinations, the actual findings concerning the design and effects of items in state wide exit examinations are remarkably sparse. In Germany, little is known about how examination items are designed (Bolle-Bovier, 1994; Brockhage and Weghöft, 1994; Kirsch, 2003). Kühn (2010) shows that science items in the state wide examinations of three selected German *Bundesländer* contain only a few experimental and context-oriented items. Altogether, the research concerning examination items in state wide exit examinations in Germany can offer only limited findings so far. The same is true for the research on examination items in other states. Currently, most analyses focus only few and very specific configurative aspects, especially in connection with Bloom's Revised Taxonomy of Cognitive Objectives as theoretical framework (Dudley, 1977; Heyneman and Fägerling, 1988; Tikkannen, 2010). A systematic international comparison of examination items focusing on several different item features is still pending. Consequently, the goal of this study is to shed light on this need.

Research Questions

This PhD-study aims to investigate and analyze the national 'status quo' as well as recent changes in the elements below for each of the selected states. The results shall explicate similarities and differences across the states and in that display common trends and state-specific features in the construction of examination items. In this context, the leading research questions (RQ) are:

(RQ 1) Which are the distinctive features of biology exam items in the selected states?

(RQ 2) Of what quality are these items, also regarding discussions about 'innovative' biology examination items?

(RQ 3) Are there common trends across states regarding the distinctive characteristics of biology examination items?

Choice of States

The study is following an international comparative study of multiple cases (Bradburn and Gilford, 1990). It has an exploratory and descriptive character since there is only little known about the design of items in the context of state wide exit examinations so far. The analysis comprises the examinations in six European states, which were chosen using the following criteria. Only European OECD states were considered to ensure a comparable economic situation, and the chosen states all had good or considerably improved performance in Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) as an indicator for 'good practice' with regard to science education. The subject biology was chosen because research in Germany reveals that biology is considered to be the most popular science subject among teachers and students (Baumert and Köller, 2000). Examination statistics show that in a number of other states biology is more popular as an examination subject than the other two science subjects (physics and chemistry).

Accordingly, we decided to analyze biology examination items of the following six European OECD states: England, Finland, France, the Netherlands, Poland and Scotland. In each state the subject biology is either a semi-compulsory or optional examination in the state wide exit examinations at basic or advanced level. The analysis is based on biology examination items and comprises examination question papers of the years 2000/1 or 2005 and 2010. Poland revised the state wide exit examinations in 2005 comprehensively. Scotland changed the biology course system in 2000 but did not immediately develop new examination guidelines. In these cases, 2001 (Scotland) and 2005 (Poland) were chosen as first measurement points.

Instrument

The item analysis shall be conducted by using a concept-oriented rating manual. The criteria for the manual are developed in a deductive and inductive way. A literature review was used to identify main criteria from existing concept-oriented rating manuals which were developed and used for item analyses (Fischer and Draxler, 2001; Blömeke, *et al.*, 2006; Jordan, *et al.*, 2006; Jatzwauk, 2007; Kulgemeyer, 2009; Kühn, 2010; Stawitz, 2010). They were adapted, supplemented and in parts condensed for the international comparison. Aspects of instruction in biology, especially with regard to innovative item characteristics (e.g. contextualization, assessment of several scientific competencies) in the sense of scientific literacy (Bybee, 1997; Roberts, 2007), can be considered in the manual. It covers the areas of format, content and cognitive operations, which each have several sub-categories (answer format; item types; content; curricular validity; contextualization; mathematical requirements; modes of representation; scientific competencies; replication; application; transfer).

Unit of Analysis

Based on Jatzwauk (2007), we decided to consider every *independent and content-based question or prompt to think or to do something* as unit of analysis (e.g. “*Explain why...*” or “*What is...*” or “*After mentioning the...*”). This approach is chosen because single prompts seem to have a better grain size than the whole item. This is particularly important because in some cases, there is more than one independent and content-based prompt within each item. Against this background, a set of regulations is developed for the identification of all units of analysis within the sample of exam items. The regulations are validated by experts and contain examples of use. To check the reliability of the regulations, all relevant units of analysis within a random sample of 20% of the examination papers were identified by two independent observers (accordance was 98.78%; since we identified the units of analysis it was not appropriate to use Cohen’s Kappa). As a result, 416 units of analysis were identified, which can be extended to 2019 units of analysis after completing the identification process of the remaining sample.

First Results

First qualitative explorations already display obvious differences between the applied item formats. While France and Finland only use open extended-response questions (both 2000 and 2010), England, the Netherlands, Poland and Scotland additionally use multiple choice and short answer questions. This is true for both 2000/2005 and 2010.

However, the main survey, which has been launched recently, will give more evidence concerning the differences and similarities between the states in all aspects and common trends regarding consideration and design of characteristics of items.

Significance of the study

From a theoretical point of view, the study can produce statements about how trends and the current practice of biology examinations items in European OECD states reflect “good” practice. The study will also deliver a frame within which e.g. the German (Kühn, 2010) as well as other examinations can be assessed regarding both their design and utilization as innovative policy tools.

From a practical perspective, the study allows for the identification of international trends concerning biology examination items at the end of general upper secondary education. In addition, the project aims to contribute to the further development of exit items.

References

- Altrichter, H., Brüsemeister, T. and Wissinger, J. (eds.) (2007) *Educational Governance: Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Baumert, J. and Köller, O. (2000) Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos and R. Lehmann (eds.), *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie-mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2* (pp. 271-316). Opladen: Leske + Budrich.
- Baumert, J. and Watermann, R. (2000) Standardisierung durch die Abiturprüfung: Zentralabitur oder dezentrale Prüfungsorganisation? In J. Baumert, W. Bos and R. Lehmann (eds.), *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie-mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Bd. 2* (pp. 341-351). Opladen: Leske & Budrich.
- Bishop, J.H. (1998) The effect of curriculum-based external exit exams on student achievement. *Journal of Economic Education*, 29, 172-182.
- Blömeke, S., Risse, J., Müller, Chr., Eichler, D. and Schulz, W. (2006) Analyse der Qualität von Aufgaben aus didaktischer und fachlicher Sicht. Ein allgemeines Modell und seine exemplarische Umsetzung im Unterrichtsfach Mathematik. *Unterrichtswissenschaft*, 34(4), 330-357.
- Bolle-Bovier, H.-P. (1994) Schriftliche Abiturprüfung im Fach Deutsch – zentrale oder dezentrale Aufgabenstellung? Ein Vergleich zwischen Aufgabenstellungen in Niedersachsen und Bayern. *Mitteilungen des Deutschen Germanisten-Verbandes* 41(4), 45-52.
- Bos, W. and Tarnai, C. (1999) Content analysis in empirical social research. *International Journal of Educational Research*, 31, 659-671.
- Bradburn, N. M. and Gilford, D. M. (1990) *A framework and principles for international comparative studies in education*. Washington, DC: National Academy Press.
- Brockhage, R. and Weghöft, R. (eds.) (1994) Zentralabitur und Naturwissenschaften. *Gymnasium in Niedersachsen*, 26(4-5), 80-82.

- Bybee, (1997) Toward an Understanding of Scientific Literacy. In W. Gräber, and C. Bolte, (eds.), *Scientific Literacy*. (pp. 37-68). Kiel: IPN.
- Dudley, B. (1977) The Mathematics of School Biology Examination Papers. *Journal of Biological Education*, 11, 41-48.
- Fischer, H. E. and Draxler, D. (2001) Aufgaben und naturwissenschaftlicher Unterricht. *Der mathematisch-naturwissenschaftliche Unterricht*, 54(7), 388-393.
- Heyneman, S. P. and Fägerling, I. (1988) *University Examinations and Standardized Testing: Principles, Experience, and Policy Options*. Washington, DC. World Bank technical paper, 78.
- Jatzwauk, P. (2007) *Aufgaben im Biologieunterricht. Eine Analyse der Merkmale und des didaktisch-methodischen Einsatzes von Aufgaben im Biologieunterricht*. Logos-Verlag.
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M., Löwen, K., Brunner, M. and Kunter, M. (2006) *Klassifikationsschema für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kirsch, W. (2003) Abituraufgaben auf dem Prüfstand – Chemie. Ergebnisse und Konsequenzen eines Erfahrungsaustauschs zwischen den Ländern – kontextorientierter Chemieunterricht als Basis für eine neue Aufgabenkultur. *Schulverwaltung, Ausgabe Hessen, Rheinland-Pfalz und Saarland*, 7(10), 267-268.
- Klein, E. D., Kühn, S. M., Ackeren, I. van and Block, R. (2009) Wie zentral sind zentrale Prüfungen? Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich. *Zeitschrift für Pädagogik*, 55(4), 596-621.
- Kulgemeyer, Chr. (2009) *PISA-Aufgaben im Vergleich. Strukturanalyse der Naturwissenschaftsitems aus den PISA-Durchläufen 2000 bis 2006*. Norderstedt: Books on Demand GmbH.
- Kühn, S. M. (2010) *Steuerung und Innovation durch Abschlussprüfungen?* Wiesbaden: VS Verl. für Sozialwissenschaften.
- Maag Merki, K. (2010) Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter and K. Maag Merki (eds.), *Handbuch neue Steuerung im Schulsystem* (pp. 145-169). Wiesbaden: VS Verl. für Sozialwissenschaften.
- Maag Merki, K., Holmeier, M., Jäger, D. and Oerke, B. (2010) Die Effekte der Einführung zentraler Abiturprüfungen auf die Unterrichtsgestaltung in Leistungskursen in der gymnasialen Oberstufe. *Unterrichtswissenschaft*, 2, 173-192.
- Roberts, D. A. (2007) Scientific Literacy / Science literacy. In S.K. Abell and N.G. Lederman (Eds.), *Handbook of research on science education* (pp. 729-780). Mahwah, NJ: Lawrence Erlbaum Associates.
- Stawitz, H. (2010) *Auswirkungen unterschiedlicher Aufgabenprofile auf die Schülerleistung*. Berlin: Logos Verlag.

Vogler, K. E. and Carnes, G. N. (2009) *Comparing the Impact of a High School Exit Examination in Science Teachers' Instructional Practice*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, California.

Tikkanen, G. (2010) *Chemistry matriculation examination as summative assessment tool*. Doctoral Dissertation. University of Helsinki.