Paper presented at the International Test Commission Conference
Vancouver, Canada; July 1-4, 2016

Standard Lists of Creative Answers Leave out Correct Solutions: An Example from PISA 2012

Alexander Poddiakov

Department of Psychology
National research University Higher School of Economics
E-mail: apoddiakov(at)hse.ru

Abstract

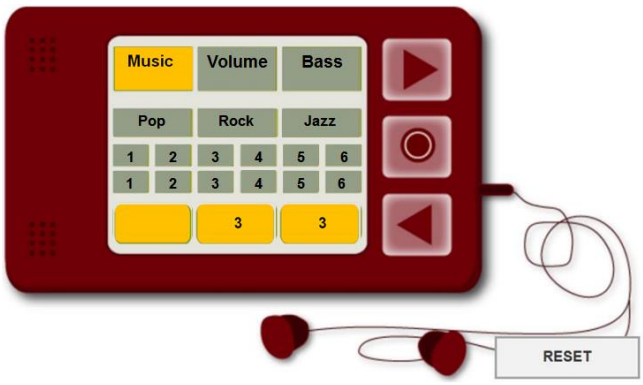
Interactive complex problem solving is an important 21st century skill (Neubert et al., 2015). Items of the PISA 2012 included interactive problems requiring exploration of a novel device (e.g., a virtual MP3-Player) and ideas for its modifications. The aim of our analysis is to show that the use of "standard lists of creative answers" used to assess creative problem solving can lead to missing some correct solutions. The object of the analysis is the fourth question of the MP3-Player Unit: "Describe how you could change the way the MP3 player works so that there is no need to have the bottom button" (http://www.oecd.org/pisa/keyfindings/PISA-2012-results-volume-V.pdf). Authors of the item write: "There is no single correct answer, and students may think creatively in devising a solution". Nonetheless, they limit the number of correct answers to the six that are described in the guidelines. Other responses are not accepted and result in a score of 0 (PISA 2012; https://nces.ed.gov/surveys/pisa/doc/cb-mp3.doc). We argue that the authors' list of correct answers is not complete: it includes neither double successive nor simultaneous clicks, even though "double-clicking" is a common ergonomic solution in modern devices. A possible paradoxical reason is that the authors agree entirely with their example of universal rules formulated about 15 years ago, before the era of multi-touch devices: "at no point two buttons have to be pressed simultaneously" (Greiff, 2012, p. 52). However, one can easily show that solutions based on different successive and simultaneous double-clicks should be included in the list of correct answers for the item (Poddiakov, 2012). The paradigm of assessing creative answers in accordance with lists of criteria prepared in advance ("standard lists of creative answers") seems to be an oxymoron in the context of assessing 21st century skills and needs to be changed. (http://conferences.educ.ubc.ca/index.php/itc2016/itc2016/paper/view/384)

In 2012 the participants in the Programme for International Student Assessment (PISA) were offered new types of problems, i.e. interactive problems, for the first time in the history of large-scale testing and assessment.

The key feature of these problems is that they require a student to practically explore a new complex technical system with properties not known in advance.

For example, a student had to experiment with a new MP3 player without any instructions. The aim was to reveal the internal hidden relationships between pressing different buttons and the modes of work of the device and, based on this, to answer test questions.

Including interactive problems into international assessment seems both crucial and timely. Exploration of novelty and experimenting with new systems belong to key skills of the 21st century, and this should be reflected in the assessment of skills.



Introductory part of the item: "A friend gives you an MP3 player that you can use for playing and storing music. You can change the type of music, and increase or decrease the volume and the bass level by clicking the three buttons on the player (◁, ⊙, ▷)" [*PISA 2012..*, 2013, p. 127. http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2012-assessment-and-analytical-framework/problem-solving-framework_9789264190511-6-en#page5]

Minimally complex items

The authors of the interactive items in PISA call their items "minimally complex items". What are their features? Why are they called complex?

Let's look at their structure.

For example, the player has 4 control buttons and more than 20 windows. It means that the function of each button is not constant, but depends on the state of the device and on the sequence in which the buttons were pressed before. Four buttons control a multitude of outputs via a complex net of connections. So this makes the item complex.

Why are these items *minimally* complex?

This is because really complex computer scenarios designed in this approach not for PISA but for cognitive studies include thousands of mutually related variables in dynamic systems of governing a virtual country or city. The founder of this experimental approach to complex problem solving (or, in wider terms, to complex cognition) was Dietrich Dörner. He pioneered these studies in the 1970s. His complex scenarios (Moro, Lohausen, etc. – see [Dörner, 1997]) preceded famous computer strategic games such as SimCity, Civilization, etc.

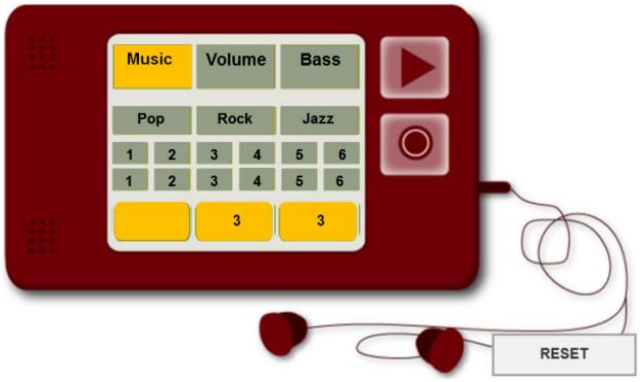
Assessment of creativity in PISA 2012

Exploration and mastery of novelty are always challenging and require creative thought. So it seems natural that designers of interactive items for PISA set an aim to assess how students can generate creative solutions working with these complex objects. Not all items to do with the interactive objects were creative, most of them were actually multiple choice. However one item was creative.

And this is a test item I see serious flaws with. I'll consider the case of the PISA creative item and show that there is a serious mistake in its scoring code. And this concrete mistake in the scoring code is made because of a more general methodological mistake of principle. This methodological mistake can be called "standard list of creative answers".

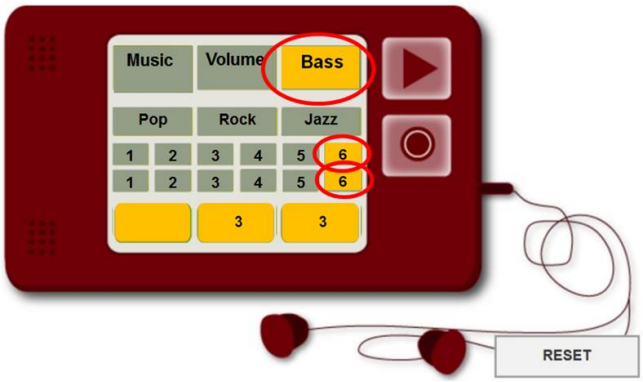
The creative item is the following: "Describe how you could change the way the MP3 player works so that there is no need to have the button ◁. You must still be able to change the type of music, and increase or decrease the volume and the bass level".

Thus, the device should look like this.



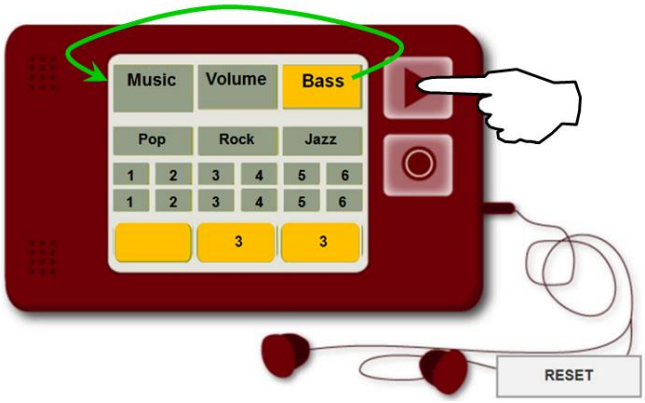
What's the problem for the participants?

It is impossible to go back from any of the extreme right positions without the bottom button.



What to do?

For example, as the designers write, to "use the one arrow to go all the way round (in a circle)" (https://nces.ed.gov/surveys/pisa/doc/cb-mp3.doc)



The authors write: "There is no single correct answer, and students may think creatively in devising a solution. <…> In the field trial, this was by far the hardest item in the unit, with only 25% of students gaining credit" [*PISA 2012.., 2013*, p. 133].

Let's look at the screenshot with scoring codes in more detail. There is the list of correct answers getting full credit (I have circled them in green) and the list with examples of the other answers not getting credit at all (I have circled them in red).

Full Credit

Code 1: Gives an answer that describes how the MP3 player could still operate with only one arrow button
- Change the way the top button works so that once you reach the right side of the display, one more click takes you back to the left of the display
- Using one arrow, each line cycles around e.g. Music-Volume-Bass-Music
- The right arrow could just take you back to the far left of the screen if you reach the rightmost entry – for example, once you are on "bass", pushing the right arrow button could take you back to "Music"

- The volume is set at 3 by default. If you want to change it to two or one, it could be set up so that when you click the middle button to set the volume, it defaults to one (the lowest setting). Then you can use the right arrow to change it to whatever you want.

- When you want to change a property and you move to the line it is on, it should default to the lowest setting for that property.
- Use the one arrow to go all the way round (in a circle). *[Minimal.]*

No Credit

Code 0: Other responses
- It would work without that button.
- You could change it so it didn't need that button. *[No explanation.]*
- The middle button could move you to the left. *[Insufficient explanation.]*

Code 9: Missing

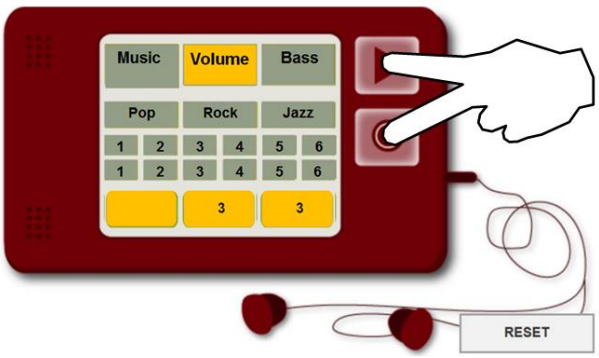
The screenshot with the scoring codes for the creative item [http://www.umac.mo/fed/pisa/research reports/20141229_MS12digital_eng.pdf, p. 88]

Thus, in spite of the fact that the authors of the item write: "There is no single correct answer, and students may think creatively in devising a solution", they limit the number of correct answers to the six or seven that are described in the guidelines. Other responses are not accepted and result in a score of 0.

Yet one can easily show that the list of correct answers is not complete.

Which are the right solutions that have been left out?

The list does not include *simultaneous clicks* on two buttons, even though simultaneous clicking is a common ergonomic solution in modern devices.

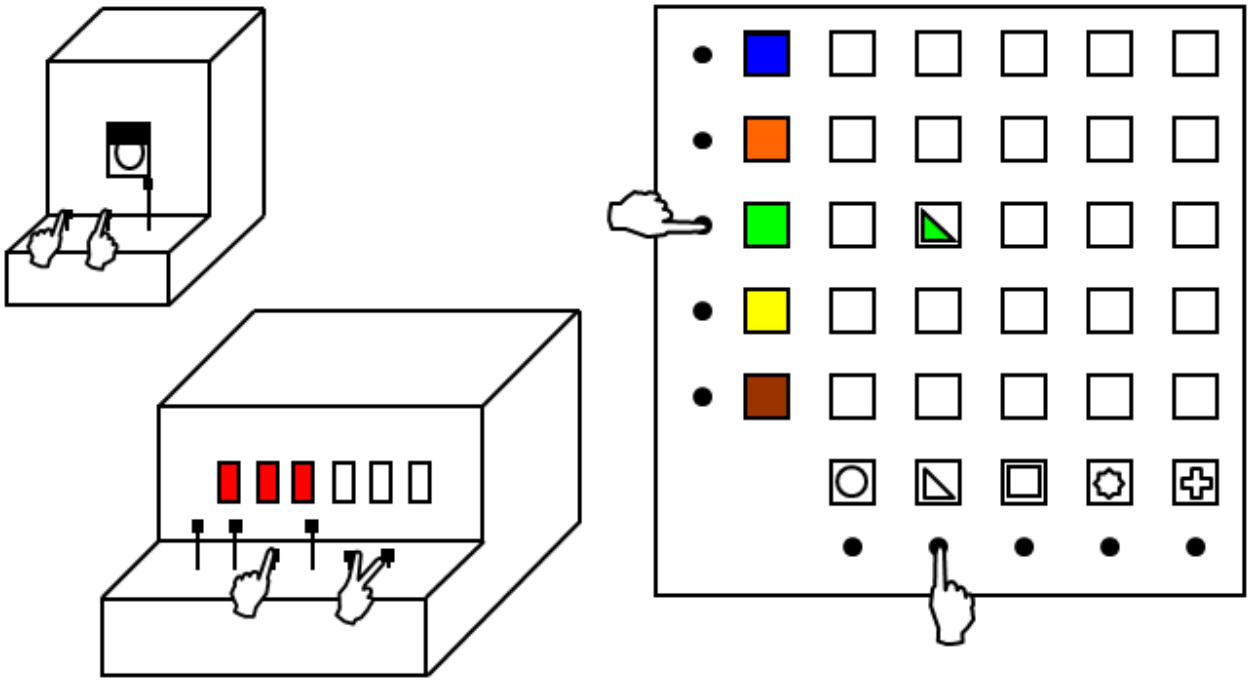


A possible paradoxical reason is that the authors agree entirely with their example of universal rules formulated about 15 years ago, before the era of multi-touch devices: “At no point two buttons have to be pressed simultaneously” [Vollmeyer, Burns, 1999; quoted from Greiff, 2012, p. 52].

Can children and teenagers invent simultaneous clicks?

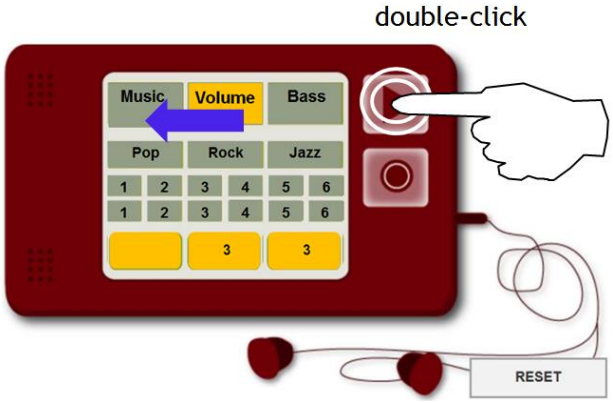
Since the middle of the 1980s I have been studying children’s independent experimentation with problem solving toys. The toys had chord keyboards which required simultaneous pressings. A participant had to understand that such pressings are necessary to make an object react in the desired way and invent different simultaneous pressings. Preschool children and secondary school students could successfully experiment with such devices and found a lot of different combinations. One can see in the photo the boy is pressing simultaneously 3 buttons with his right hand and at the same time is switching the toggle switch with his left hand.

I used different devices, and the participants coped with them (Poddiakov, 2011).



It was easy for many children to come up with the idea of simultaneous pressings. Yet simultaneous clicks are not included in the list of correct answers.

The second solution left out of the list of correct answers is double-clicking. Double-click on the top button or on the middle button can return the mode to the value at the extreme left position.



Double-click is a possible ergonomic solution in modern devices too. So the list of the correct answers rather reflects skills not of the 21st century, but of the 20th century.

An indefinite number of solutions to the inventive problem

Moreover, one can mathematically prove that the problem to find ways to do without the bottom button has an indefinite number of solutions. Some of them are rather evident like simultaneous pressing or double-clicks. Some of the other ones can be sophisticated like a secret code for a safe. However, in general any combination (or sequence) of pressings of one or two top buttons can code the function of the third button which is now absent. Choices of the simplest ones or sophisticated ones can depend on the situation. The player is a model of a finite automaton and this gives the opportunity to create an indefinite number of arbitrary combinations and sequences.

If we consider real-life inventive activity, we see that a lot of inventions in different domains were made when the list of all the possible solutions seemed exhausted.

So the paradigm of assessing creative answers in accordance with lists of criteria prepared in advance ("standard lists of creative answers") seems to be an oxymoron in the context of assessing 21st century skills and needs to be changed. Truly original solutions have a way of not being on the list of already known answers, but of enlarging it.

I think that the last question to this creative item could be like this: "If anyone invents a lot of solutions how to do without the bottom button, can s/he be sure that the list of solutions is exhausted and nobody can invent at least one more solution absent in the list?"

Also it would have been more accurate to include the following phrase in the guidelines: "We understand that any exhaustive, prepared in advance, list of solutions for the inventive problem is impossible, but here, in this field trial study, as a palliative, we limit a number of correct answers to 6 (or 7, or 17, etc.)". Otherwise teachers, school psychologists, administrators and parents interested in PISA can consider the recent list of correct answers and the instruction to estimate all the other answers as 0, as the very model of assessment of creative solutions. Yet this runs counter to the very idea of creativity.

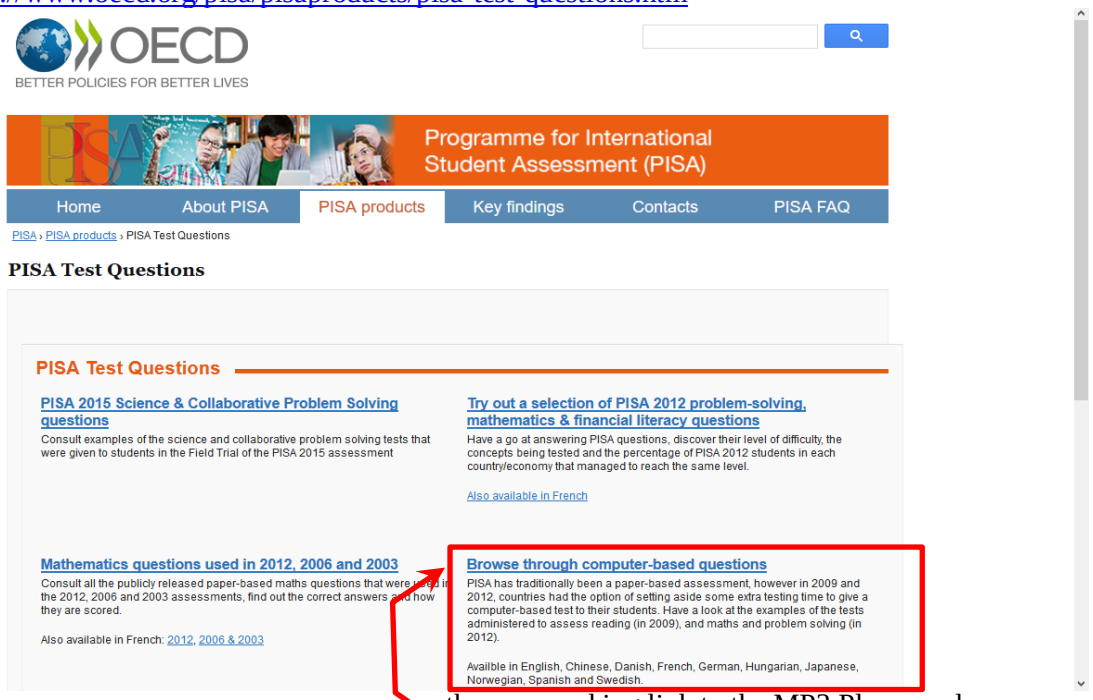
What about PISA items of 2015?

In the document *PISA 2015 Released Field Trial Cognitive Items* one cannot find words "creative" or "creativity" at all [http://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf]. I think that this could mean that PISA test designers have become aware of the problems associated with assessing creative tasks.

The disabled links to the MP3 Player and some other interactive items of PISA 2012

Concerning the interactive items of 2012, now the link leading to them, including the item with the player, is no longer live on the PISA site. It was working for a few years, but it seems to have been disabled in 2016.

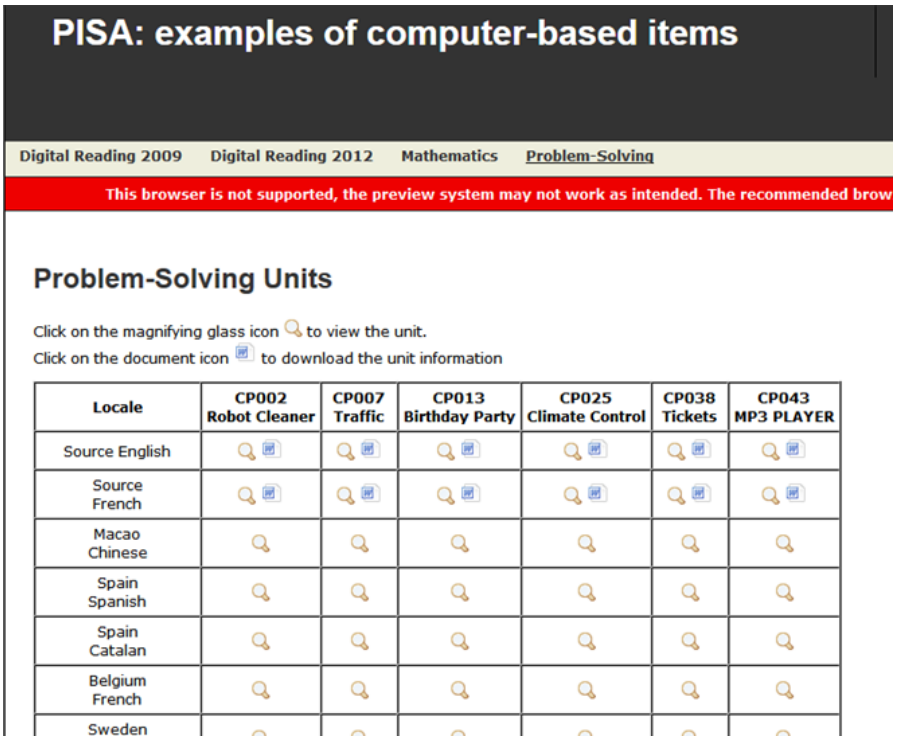
http://www.oecd.org/pisa/pisaproducts/pisa-test-questions.htm



the non-working link to the MP3 Player and some other interactive items on the PISA site

Now, only with the help of archive web machines (like http://wayback.archive.org) can you go back and see how the first page looked. Nobody can play with interactive items such as *Robot Cleaner* and *MP3 Player* and check them now. You can play with some other items but not with these ones. Yet one can work with PISA materials in which the items and scoring codes are described.

http://wayback.archive.org/web/20151027005647/http://erasq.acer.edu.au/index.php?cmd=toProblemSolving



Conclusion

Let me emphasize again that introduction of interactive problems in assessment, including international assessment, is timely and absolutely necessary. Exploration of novelty and experimenting with new systems belong to key skills of the 21st century, and these skills must be assessed. However experimenting with novel systems requires creativity from the participants. How can we approach it? Standard lists of creative answers do not seem an appropriate measuring tool for it. So what is an appropriate solution going forward?

Mass assessment of the creativity of hundreds of thousands of participants requires a standard, doesn't it? If so, one can only say that this standard should be very non-standard.

More detailed analysis of interactive items and paradoxes of mass assessment of creativity is given in the article (Poddiakov, 2012).

References

Dörner, D. (1997). *The logic of failure: Recognizing and avoiding error in complex situations*. New York: A Merloyd Lawrence Book.

Greiff, S. (2012). Assessment and Theory in Complex Problem Solving. A Continuing Contradiction? *Journal of Educational and Developmental Psychology, 2*(1), 49-56.

Neubert, J. C., Mainert, J., Kretzschmar, A., & Greiff, S. (2015). The Assessment of 21st Century Skills in Industrial and Organizational Psychology: Complex and Collaborative Problem Solving. *Industrial and Organizational Psychology*, 8(2), 238-268. doi:10.1017/iop.2015.14.

PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy. OECD: OECD Publishing, 2013. http://dx.doi.org/10.1787/9789264190511-en.

PISA 2015 Released Field Trial Cognitive Items. http://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf.

Poddiakov, A. (2011). Didactic Objects for Development of Young Children's Combinatorial Experimentation and Causal-Experimental Thought. *International Journal of Early Years Education, 19*(1), 65-78. doi:10.1080/09669760.2011.571001.

Poddiakov, A. (2012). Complex Problem Solving at PISA 2012 and PISA 2015: Interaction with Complex Reality. *Obrazovatel'naya Politika, 6,* 34-53. http://ssrn.com/abstract=2432598.

Vollmeyer, R., & Burns, B. D. (1999). Problemlösen und Hypothesentesten [Problem Solving and Hypothesis Testing]. In H. Gruber, W. Mack & A. Ziegler (Eds.), *Wissen und Denken. Beiträge aus Problemlösepsychologie und Wissenspsychologie* (pp. 101-118). Wiesbaden: Deutscher Universitäts Verlag.