

The notion of authenticity in the PISA units in physical science: an empirical analysis

Laura Weiss · Andreas Müller

Received: 26 February 2014 / Accepted: 6 February 2015 / Published online: 8 March 2015

© Gesellschaft für Didaktik der Physik und Chemie (GDPC); Fachsektion Didaktik der Biologie im VBIO (FDdB im VBIO) and Springer-Verlag GmbH Berlin Heidelberg 2015

Abstract The notion of “authenticity”—in the sense of a close relationship to actual and real (or at least realistic) contexts ensuring “relevance to students’ interests and lives”—is essential to PISA’s understanding of scientific literacy (OECD, 2007, p. 36), as well as for tradition of Context-Based Science Education (CBSE; see e.g. Bennett et al., 2007). Looking at the PISA units, there can be no doubt about their factual authenticity, i.e. that they are indeed based on real-life contexts as stated above. The assumption that these situations are perceived *by pupils* as interesting and authentic (personal authenticity) is more arguable. It therefore appears necessary to investigate this perception, as it is supposed to be essential both for motivation and cognitive activation of learners.

This contribution thus presents a study about the perception of the reality connection/authenticity (RA) as well as of interest and engagement (IE) related to the PISA physical science units, with a triple perspective: pupils’ and teachers’ perceptions, and additionally teachers’ assumptions about pupils’ perceptions. Results show that, for the available PISA items, pupils’ perceived authenticity and

interest are at best medium, contrary to the basic assumption of PISA.

Moreover, the present approach allows to validate and to quantify the common wisdom that there is a large gap between the perceptions of teachers and pupils (RA: Cohen $d=1.34$; IE: $d=1.60$; $p<0.001$ for both). When asked for their assumptions about pupils’ perceptions, teaching experts adjust their own perceptions to lower values, but still show considerable overestimation of pupils’ interest and perceived authenticity (RA: $d=0.66$, $p<0.05$; IE: $d=0.85$, $p<0.01$). We conclude with a discussion of some limitations of the study and several implications for classroom practice of these findings.

Keywords Authenticity · Context-based science education · PISA science · Pupils’ perceptions · Teachers’ perceptions

Das Verständnis von Authentizität in den PISA-Testeinheiten zur Physik: Eine empirische Analyse

Zusammenfassung Der Begriff der Authentizität im Sinne einer engen Verbindung zu realen (oder mindestens realistischen) Kontexten, die „relevance to students’ interests and lives“ (OECD, PISA 2006. Science competencies for tomorrow’s world, volume 1: Analysis, p. 36, 2007) sicherstellen sollen, ist grundlegend für PISA’s Verständnis von scientific literacy, und gleichermaßen für die Tradition der „Context-Based Science Education“ (CBSE; see e.g. Bennett et al., Science Education 91:347–370, 2007). Mit Blick auf die PISA Units1 kann kein Zweifel an ihrer faktischen Authentizität bestehen, d. h. dass sie in der Tat auf lebensweltlichen Fragen und Situationen im og. Sinne basieren. Dass diese Situationen auch durch *Schüler*

A note on terminology: Following PISA, we will use the term “unit”, as most of the PISA problems are clusters of several items (and not a single “item”). Moreover, we will refer to “released units” for those made available for the public (8 out of 36 units; OECD 2007).

A. Müller (✉) · L. Weiss
Institut Universitaire de Formation des Enseignants (IUFE),
Université de Genève,
Pavillon Mail, 40, Boulevard du Pont d’Arve,
1205 Genève, Switzerland
e-mail: Andreas.mueller@unige.ch

L. Weiss
e-mail: Laura.Weiss@unige.ch

als interessant und authentisch (persönliche Authentizität) wahrgenommen werden, ist hingegen stärker fraglich. Es erscheint somit notwendig, diese Wahrnehmung zu untersuchen, als sie als Grundlage sowohl von Motivation wie kognitiver Aktivierung angenommen wird.

Dieser Beitrag stellt deshalb eine Untersuchung zu „Realitätsbezug/Authentizität“ (RA) sowie „Interesse/Engagement“ (IE) mit Bezug auf die PISA Units (physical sciences) unter einer dreifachen Perspektive vor: Die Wahrnehmung (Einschätzung) durch Schüler und Lehrer, sowie die Annahmen der Lehrer über die Schülereinschätzung. Die Ergebnisse zeigen für die veröffentlichten PISA Units eine höchstens mittlere Einschätzung von Authentizität und Interesse durch die Schüler, im Gegensatz zu der Ausgangsannahme von PISA. Außerdem wurde ein deutlicher Unterschied zwischen Lehrer- und Schülereinschätzungen gefunden (RA: Cohen $d=1,34$; IE: $d=1,60$; $p<0,001$ für beide). Werden Lehrer nach ihren Annahmen über die Schülereinschätzung befragt, adjustieren sie ihre eigene Einschätzung nach unten, überschätzen aber immer noch erheblich, wie authentisch und interessant Schüler die Units finden (RA: $d=0,66$, $p<0,05$; IE: $d=0,85$, $p<0,01$). Der Beitrag schließt mit einer Diskussion von einigen Einschränkungen der Studie, sowie von Konsequenzen dieser Ergebnisse für den Unterricht.

Schlüsselwörter Authentizität · Context-Based Science Education · PISA Naturwissenschaften · Schülerwahrnehmungen · Lehrerwahrnehmungen

Two student quotes to the question: “What is Boring” (in School):

“Physics. I have never, nor will I ever, see the point in or understand physics.

It always seemed pointless spending hours of experimental time proving what was already proven, or that black wasn't a colour, or whatever.”

“Chemistry. Learning how chemicals are used in industry is very boring

– Chemicals in the body and used in drugs are more interesting and relevant.”

From Jenkins (2006)

Introduction: Background and Rationale for the Study

A quite widespread, basic understanding of authentic learning (starting with the word origin: gr. *authentikós* “true”; lat. *authenticus* “reliable”) is that it should be related to actual, real(istic), genuine contexts and experiences learners are supposed to encounter. This point of view is also essential to and strongly advocated by PISA (OECD 2006, 2007), as underlined e.g. by the in-depth analysis of Fensham (2009): “real world contexts have [...] been a central feature of the OECD’s PISA project for the assessment of scientific lit-

eracy among young people“. Moreover, this is also the basic understanding and “common numerator” of the variety of approaches addressed as “context based science learning” (CBSL; Bennett et al. 2007).

The goal of scientific literacy as understood by this current of science education should also contribute to develop and strengthen interest for the sciences and technology, in particular to counteract the widespread disaffection towards these areas, a current problem in the western countries (Rocard et al. 2007). As this problem is particularly pronounced for the physical sciences (Bøe et al. 2011; Murphy and Whitelegg 2006; Jenkins 2006; Zwick and Renn 2000), we focus on this domain in the following.

With this background, it is a central issue for PISA that its units are authentic and motivating for young people and it is this point of view that we analyze in this article, based on empirical data of a survey with secondary level I pupils and teachers in Geneva. The contribution is an extension of our preceding studies aimed to better understand and qualify what PISA actually evaluates: a first paper on the comparison between the PISA science units and the science curriculums of French speaking Switzerland (Weiss 2010) and a second one where the compatibility between the Inquiry Based Teaching (IBT) and the science PISA survey was discussed (Weiss, attempted for publication).

The paper is organized as follows: after giving theoretical background about the notion of authenticity, we describe the PISA choices for its questions in this respect. We then proceed with a description of our choice of the PISA units for the survey, of the sample, and the instruments of the study. Results from the pupils’ and teachers’ sample about their perception of these PISA units will be discussed and compared between themselves, including some limitations of the study, and finally several conclusions concerning classroom implications and future research will be discussed.

The Question of Authenticity and Interest

An understanding of “authentic contexts” as being related to actual, realistic and genuine contexts and experiences learners are supposed to encounter is basic, but far from trivial or educationally shallow (even if within Context-Based Science Education (CBSE) more far reaching conceptualisations exist, see below). PISA states two important points about that (OECD 2006): First, such problems, to be encountered in real-world settings (“factual authenticity”), are usually not stated in the disciplinary terms to be learned or applied. Thus, a work of “translation” with terminological and conceptual reframing has to be carried out, representing a first step of cognitive activation. Second, the disciplinary content involved is genuinely directed to solving the problem (OECD 2006, p. 81), i.e. learners can perceive that there is

Table 1 PISA 2006 science contexts (OECD 2007, p. 36)

PISA 2006 science context	Personal (Self, family and peer groups)	Social (The community)	Global (Life across the world)
“Health”	Maintenance of health, accidents, nutrition	Control of disease, social transmission, food, choices, community health	Epidemics, spread of infectious diseases
“Natural resources”	Personal consumption of materials and energy	Maintenance of human populations, quality of life, security, production and distribution of food, energy supply	Renewable and non-renewable, natural systems, population growth, sustainable use of species
“Environment”	Environmentally friendly behaviour, use and disposal of materials	Population distribution, disposal of waste, environmental impact, local weather	Biodiversity, ecological sustainability, control of pollution, production and loss of soil
“Hazard”	Natural and human-induced, decisions about housing	Rapid changes (earthquakes, severe weather), slow and progressive changes (coastal erosion, sedimentation), risk assessment	Climate change, impact of modern warfare
“Frontiers of science and technology”	Interest in science’s explanations of natural phenomena, science-based hobbies, sport and leisure, music and personal technology	New material, devices and process, genetic modification, transport	Extinction of species, exploration of space, origin and structure of the universe

a real-world problem for the solution of which some content of science or math is necessary (“problem authenticity”), instead of the problem being just an (invented, artificial) occasion to practise this content. Moreover, the combination of these two features of authenticity is also supposed to be closely linked to the (science related) self-concept, as it should be supported by the experience of actually being able to solve real-world problems using the knowledge and competences one has acquired (OECD 2006; Hattie 2009, p. 46). Summing up, “conceptual translation” and “genuine content-problem link” in this sense can justly be considered important components of scientific literacy, as PISA does. Moreover, beyond cognitive features, authentic contexts are supposed to foster attitudinal and affective aspects, in particular interest in science. Fensham (2009) states “Real world contexts from the students’ lives outside of school have the potential to generate personal intrinsic interest, and their social or global significance can add to this potential an extrinsic quality to this interest.” CBSL in general (Bennett et al. 2007), makes the same claim about the potential of linking science education to pupil’s life. There are other and partially more far-reaching understandings of authenticity (CTGV 1990; Mims 2003; Herrington and Herrington 2006). A quite comprehensive conceptualisation was given by Shaffer and Resnick (1999), who distinguish four aspects: learning (1) “related to the real-world outside school” (“factual authenticity”), (2) “personally meaningful to the learner” (“personal authenticity”), (3) providing “an opportunity to think in the modes of a particular discipline” (“cognitive authenticity”, see also Herrington and Herrington 2006) and (4) assessment in line with the learning process (“assessment authenticity”). Note, that these aspects can be refined further. For example, (3) can be understood as genuine usage of disciplinary content to solve a real-life question, i.e. the “problem authenticity” already mentioned,

or as participating in the “community of practice” of the given discipline.¹

There is a considerable body of international literature on these “authentic contexts” (see e.g. the thorough analyses of Muckenfuß (1996) or of Cariou (2010)), a review of which is beyond the scope of this contribution. We do want, however, to use the conceptual framework given so far in order to give a clear localization of the conceptualization according to PISA, which is at the focus of the present article.

PISA designers have taken account of authenticity in aligning the test units with five broad areas of “personal, social and global settings” with essential applications of science in the real life (such as health and environment), and not with the more traditional division of the science disciplines as taught in school as biology, chemistry or physics (see Table 1).

In line with the objectives sketched above, PISA’s choice of units is based on an assumption about their “relevance to students’ interests and lives, representing science-related situations that adults encounter [...] almost daily” (OECD 2007, p. 36). If the second part of this assumption is certainly true, it is not sure that the first, crucial part about the “relevance to students’ interests and lives” can be taken for granted.

Within the conceptual framework of the four authenticity aspects given above, one may state that concerning the first and third of the aspects, the approach of PISA clearly encompasses factual authenticity and problem authenticity. This becomes apparent on the basis of the description just given, and by the repeated commitment to the usage and value of tasks and problems (items) “that could be part of

¹ It makes sense to distinguish terminologically the two forms of cognitive authenticity, e.g. by “practical” and “epistemological” (or “methodological”) cognitive authenticity.

the actual experience or practice of the participant in some real-world setting”, and that PISA “places most value on tasks that could be encountered in a variety of real-world situations” (OECD 2007, p. 81). Great care was taken within the development of PISA to ensure an as good alignment of the guiding idea of science literacy and of its assessment as possible, i.e. the forth aspect of assessment authenticity is a least partially covered as well.

The second aspect, personal authenticity, is more critical to evaluate, as this is not a feature of the item alone, but of the *person* tackling with this item and her personal, i.e. subjective, perception: While the features of factual and problem authenticity can be ensured by careful work of researchers, and indeed are implemented in the PISA units, do youths of the target age group actually perceive these as actually related to *their* lives (personal authenticity)? It is this question of personal (learners) perception of authenticity, non-obvious to achieve, and non-trivial to assess, but essential to the very idea of scientific literacy, and its relation to the “external” (researchers, teachers) view, that we are mainly interested in.

The question of authenticity is of course embedded in and related to other motivational variables. Within PISA’s conceptual framework, in particular, interest in science and self-concept as science learners had been assessed (OECD 2007). Both interest and self-concept are educational objectives in their own right, and moreover, they are known to influence learning with considerable effects sizes (science interest, $d=0.68$; self concept, $d=0.43$; Hattie 2009). In the context of the present study, focusing on the personal perception about the PISA units, we also wanted to assess to which extent pupils take interest in dealing with them, and how confident they feel about doing so.

Research Questions

As explained above, our guiding question is about the personal authenticity and interest perceived by pupils in relation to the PISA units. Moreover, we are interested in teachers’ perceptions as well, and possible differences to pupil’s perceptions, i.e. we asked first for their *own* perception of the authenticity and the interest of the PISA units, the second for their judgment about *pupils’* interest and perception of authenticity.

The first guiding question is crucial in order to validate an important assumption of the PISA survey, viz. that “the contexts used for the questions were chosen in the light of relevance to students’ interests and lives” (OECD 2007). Secondly, a comparison between teachers’ and pupils’ perceptions about these issues provide important information for curricular and lesson planning, in particular when aiming at scientific literacy and context based science education.

The present contribution reports about a study about these questions, complementing the question of authenticity with other motivational variables important in PISA’s framework (science related interests and self-concept) under a triple perspective: First, pupils’ perceptions of the published PISA 2006 units related to the physical sciences; second, teachers’ perceptions of the same units as professional physicists; third, teachers’ assumptions about pupils’ perceptions. Our research questions thus read as follows:

1. “Do pupils perceive the PISA units as interesting and authentic (in the sense that they are connected to real life)?”
2. “Do teachers perceive the PISA units as interesting and authentic?” “How do teachers’ perceptions compare to pupils’ perceptions?”
3. “Do teachers have accurate assumptions about pupils’ perceptions (on the above issues)?”

Moreover, there are some complementary questions in order to relate our results to other findings, and to look whether there are any particular features of our sample. Self-concept will be assessed in order to ensure that pupils in our sample do not perceive the PISA questions as “out of reach” (within this study, this is a necessary condition to fulfill and not a research question proper). Moreover, pupils’ outcomes will be examined for a possible dependence on various characteristics (gender, grades, school levels) known to have a possible influence from previous research.

We will now turn to the methodological details on how these research questions were investigated in the study.

Materials and Methods

Selection of PISA Units

To answer empirically to the above questions, we have chosen three PISA 2006 units among the five published ones in the field of physics and chemistry; the other PISA 2006 released units are more linked to biology or earth science. The chosen units are (OECD 2007, 82 pp): *Sunscreens*, *Greenhouse* and *Clothes* (two further units linked to physical sciences are *Grand Canyon* and *Acid rain*). Our choice depends mainly on two constraints:

- The questionnaire was administered in a regular physics lesson², by the teacher of physics, thus the units had to deal with physics.
- Each PISA unit implies a considerable reading time. It was not possible to ask pupils to read the unit and answer to roughly 15 items for more than three units.

²In secondary level I schools in Geneva there are two science courses: physics (which contains a part of chemistry) and biology.

Table 2 Breakup of pupils sample (m = boy; f = girl; $n.m.$ = no mention)

Number of pupils	8th grade	9th grade	
Level A (high)	28 (15 m; 13 f)	94 (42 m; 51 f; 1 n.m.)	122 (57 m; 64 f; 1 n.m.)
Level B (low)	5 (2 m; 3 f) 33 (17 m; 16 f)	17 (15 m; 2 f) 111 (57 m; 53 f; 1 n.m.)	22 (17 m; 5 f) 144 (74 m; 69 f; 1 n.m.)

The selection of the physical science units out of five was done according to the following characteristics: *Sunscreens (Sun)* and *Acid rain* are both about an experiment done by pupils to verify a hypothesis, so we had to choose one of them. As *Acid rain* needs some knowledge on chemical reactions, which is not part of the physics curriculum of the lower secondary school and because the problem of the acid rains is no more in everyday newspapers, we preferred *Sunscreens*. *Greenhouse (Gr)* is about the CO₂ problem, which seems an important society subject, still discussed in the media in relation with the Doha climate changes meetings. Finally, we chose *Clothes (Cl)* for its originality, and let apart *Grand Canyon*, which mixes physics, earth and life sciences.

The published characteristics of the items of the three units (reported in the online material, see Table 2) show that they have a wide range of difficulty from -1.39 to $+1.93$, which is reflected by the percentage of correct answers going from 27 to 79% and by the PISA threshold (score) going from around 400 to 700.

In order to assess whether specific features of the local curriculum could diminish or enhance pupils' and teachers' perceptions of authenticity (interest, self-concept) for the selected PISA units, we first checked for the content. Comparison with the local curriculum shows, that the pupils of our sample are not supposed to have factual pre-knowledge about these units (this is in line with the PISA assessment philosophy, which does not require such factual knowledge, because the information is given in the unit basis text; OECD (2007), p. 20). On the other hand, the local curricula of physics as well as of biology emphasize that "the scientific method" has to be experienced by pupils as often as possible. They are supposed to be taught to reason in a scientific way, to understand what a hypothesis is, how evidence can validate or invalidate a hypothesis, etc. These competencies are in line with the idea of "scientific literacy" behind the PISA questions and seem important to answer correctly the items of *Sunscreens* or of *Greenhouse*. We can thus conclude that the pupils of our sample will not have specific content knowledge about the units' questions, but will not be unfamiliar with the kind of reasoning necessary to answer them.

Table 3 Sample questions for pupils**Reality connection/authenticity (RA)**

The answers to these questions are useful for everyday life

Interest/Engaging (IE)

At home I would look up in books or on the Internet, etc. to know more about the subject of this problem

Self-Concept (SC)

I always succeed well in answering to this kind of questions

Participants

Pupils' sample The questionnaire was answered by pupils in the age group from 14 to 16 years, in June 2011 during one of the last physics lessons of the year, during one period (45 min). Pupils belonged to four different lower secondary schools and to ten 9th grade classes and four 8th grade classes, distributed in ten A level (high achieving) classes and four B level (low achieving) classes³. The detailed sample breakup is given in Table 2.

Physics teachers sample This sample consists of 20 persons teaching in secondary school and/or in teachers' education, who volunteered to participate in the study. As all but two participants teach at school, we will refer to them as "teachers", for short.

Instruments and Procedures

Within this study, the following components of pupil's motivation regarding the PISA units are investigated: personal authenticity, in the sense of perceiving the connections to real life (reality connection/authenticity; RA); intrinsic interest, in the sense of liking to know about a given topic, and being ready to engage in further inquiry and learning about it (interest/engaging; IE); self-concept (SC), in the sense of considering oneself being able to learn about given topic, and to solve questions related to it (see Table 3 for sample items). These variables were assessed with instruments well established and validated in the literature. The IE and SC subscales were adapted from a large sample physics interest study ($N \approx 10000$, Hoffmann et al. 1997) within a broad research program on context based science learning (Müller 2009; Kuhn 2010; Weiss 2010; Vogt 2010). The RA subscale was developed within the same program and, together with the other subscales, thoroughly validated through several studies (total $N \approx 1500$, Kuhn 2010; Vogt 2010; Müller et al. 2011; reliabilities in these preceding studies were high to very high; α_c from 0.86 to 0.94). These

³The secondary I school in Geneve comprises three grades. Physics is taught optionally in the 8th grade and compulsory in the 9th. According to the results of the primary school, well-performing pupils (around 70%) enter the A level requested to enter academic track schools, the other are admitted in the level B (around 25%).

subscales were combined and administered as a three-component motivation⁴ instrument addressing IE, RA and SC as factors of motivation (7, 7, 10 items for IE, RA, SC, respectively; all scales 6-level Likert type); reliabilities of total and subscales were again high to very high (α_c for IE: 0.88, RA: 0.92, SC: 0.89; total: 0.90). The rationale for aggregating the subscales is to have an overall measure of motivation (in the same sense that other instruments, e.g. the Force Concept Inventory (Scott et al. 2012) also aggregate different subscales). This is consistent with scale properties obtained in the large sample validation mentioned above: one the one hand, a high reliability of the overall scale was obtained ($\alpha_c=0.93$ in Kuhn (2010); 0.9 in this study); on the other hand factor analysis led to the 3-factor structure of RA, IE and SC used in the above-mentioned and the present study (see Kuhn 2010 for details).

A French translation of this instrument was produced (the 6-level Likert scale was adapted for similarity with the school notes in Geneva, which go from 1 (worst) to 6 (best)). Its correctness and comprehensibility were validated by native speakers. It was preceded by the following instruction: “*Here are 3 units of the 2006 PISA survey on science which permitted to make a comparison about fifteen years old pupils’ scientific competencies of 60 countries. Please do not answer the items, but give your opinion about them by answering to the following questionnaire*”. So pupils had to read each PISA unit (text and items) and, without solving the items, fill in the questionnaire.

For reasons of test length, each pupil only answered to a subset of the 24 questions: all pupils answered to the RA items, and half of the class (random choice) to either the SC or the IE items (valid questionnaires: RA: 133, SC: 71, IE: 62, respectively). Individual test scores on each subscale were calculated as the percentage relative to the maximum degree of agreement. For the comparison with teachers’ perceptions about RA and IE these two subscales were aggregated into a sum scale (not SC, see below); the rationale and justification is the same as above for the original sum scale containing all three components. Below, averages and standard deviations of these measures (RA, IE, Mot = sum scale) are given for each unit as well as for the three units together (called “PISA”).

Gender, grades, and academic (school) level were included in order to see, whether the bigger interest of Swiss

boys in science found by PISA survey (Zahner Rossier and Holzer 2007, p. 23), particularly in the French speaking part of Switzerland (Moreau 2008, p. 152) and in Geneva (Moreau 2008, p. 163) will show up also in this sample, whether a better school achievement is linked to a better perception of interest and authenticity of the PISA units or a bigger science self-concept and whether older pupils are possibly more aware of the authenticity of the PISA units or more interested in the treated topics than younger pupils.

Within the pupils’ sample, the covariates just mentioned were taken into account, and results analyzed with ANOVA. For the comparison between pupils’ and teachers’ perceptions, as well as of teachers’ assumptions about pupils’ perceptions and the actual perceptions of the latter (P-T and P-TP comparison, described below), significance levels were analyzed by t-tests with Welch-Satterthwaite correction for unequal sample sizes, and after checking for prerequisites: a Kolmogorov-Smirnov test showed no significant deviations from normality for any of the samples or motivation variables ($p \geq 0.2$ in all cases), and a Levene test showed no significant deviations for variance homogeneity (RA: $p=0.4$, IE: $p=0.6$ for the P-T comparison; RA: $p=0.7$, IE: $p=0.8$ for the P-TP comparison). Finally, effect sizes were computed as Cohen d (using the standard deviation of the control, i.e. pupil’s group).

In order to compare pupils’ to teachers’ perceptions, a second, corresponding version of the questionnaire was constructed for teachers on the basis of the pupils’ questionnaire with the same six levels’ Likert scale. Teachers were asked to answer for all 5 available physical science units, but with only 8 items (instead of 24), eliminating the redundant ones, and keeping test time short (with the reliabilities for the shortened version still acceptable; α_c for RA: 0.76, IE: 0.94, total: 0.91). As this study inquires about possible differences in teachers’ and pupils’ perceptions of authenticity and intrinsic interest, the SC subdimension was left out from the teachers’ questionnaire (which moreover is arguably the most difficult aspect to be judged from outside). We asked however for all five PISA units in order to compare the motivational potential of the three units chosen for the pupil study to those not chosen (a great difference in favour of the latter would be a reason to reconsider this choice).

Items in the teachers’ questionnaire correspond to those in the pupil questionnaire and asked about the same issues under two aspects: Their own perception of the units, and their assumptions about pupils’ perception. The following sample (Table 4) shows the correspondence between the questionnaires.

As we are asking for teachers’ assumptions of pupils’ perceptions (instead of asking the perception as such), the items are not exactly the same, but up to this difference there is a one-to-one correspondence between two RA items in pupils’ questionnaire and items (1b) and (2b) in the teachers

⁴Note on terminology: PISA uses the term ‘interest’, which is related to possible *objects* of interest (topics, activities). Interest, in turn, is considered as an important component of (intrinsic) motivation, which is related to *decision and action* in general, and *learning* in particular. Self-concept is considered to be another important components of motivation (Rost 2006; entries on “Interesse” and “Intrinsische und extrinsische Motivation”), as well as, within the framework of CBSE, perception of authenticity and reality connections. We thus follow the PISA terminology and speak about science related ‘interest’, but use the term ‘motivation’ when dealing with the combined questionnaire.

Table 4 Correspondence between teachers and pupils questionnaires

Pupils: *The answers to these questions are useful for everyday life*

Teachers (own perception): *According to you, are the answers to these questions useful for everyday life?*

Teachers (assumption about pupils' perceptions): *According to you, would the pupils consider that the answers to these questions are useful for everyday life?*

questionnaire and two IE items in the pupil's questionnaire to items (3b) and (4b) in the teachers' questionnaire. Moreover we prompted teachers to give their personal opinion: *according to you* to avoid that they limit themselves to general opinions about PISA and to show them the importance of their sincere answer.

The item (3a) in the teachers' questionnaire asks whether they would use the given PISA unit for their own teaching. This question was included for two reasons: first to invoke their professional competence and second as an indicator of units questions possibly inappropriate within the Geneva curriculum. In a sense, that enhances the parallel with the pupils' questionnaire: pupils answered the questionnaire during the class, in the role of pupils; teachers in their role as experts for physics education (and neither as layman nor physicist, both not concerned by science teaching). Abbreviations (RA, IE, Mot = sum scale) are as above, and the scores are also reported as the percentages relative to the maximum degree of agreement.

Results

Pupils' Perceptions: Main Results

The results show that pupils globally do not consider the contexts of the three units as very authentic (RA around 50%), nor interesting (IE around 40%), see Table 5. The RA result for *Greenhouse* is slightly, but significantly higher ($p=0.003$) than the other units, but is not considered more interesting (IE: $p=0.97$).

These results are probably linked to the fact that pupils are aware of the problem of the atmosphere's heating, which is very present in newspapers and discussed in the society. On the other hand, *Clothes* which could be considered less

Table 5 Pupils' perceptions (averages, standard deviations; values are percentages of maximum score, see 4.3) of the three selected units ("PISA" refers to their average value) for the three subdimensions RA, IE and SC

Values in %	RA (7 items)	IE (7 items)	SC (10 items)
<i>Sunscreens</i>	47 (21)	37 (19)	52 (19)
<i>Greenhouse</i>	52 (23)	39 (22)	53 (22)
<i>Clothes</i>	46 (24)	41 (20)	51 (22)
PISA (3 units)	48 (18)	39 (18)	52 (19)

Table 6 Pupils' perceptions according to gender (F/M: female/male; averages and standard deviations; values are percentages of maximum score, see 4.3); for all differences but one, $p<0.05$ (only exception: SC for *Clothes*)

Values in %	RA		IE		SC	
	F	M	F	M	F	M
<i>Sunscreens</i>	42 (20)	51 (20)	28 (19)	43 (17)	46 (18)	57 (18)
<i>Greenhouse</i>	47 (22)	57 (23)	29 (20)	47 (21)	47 (19)	59 (23)
<i>Clothes</i>	42 (23)	50 (24)	34 (22)	48 (16)	49 (21)	53 (23)
PISA (3 units)	44 (16)	52 (18)	30 (18)	46 (15)	48 (18)	56 (20)

authentic because the scientific innovation it is about is not widely spread in medias, scores equally with *Sunscreens*, the topic of which is extensively dealt with every spring in Switzerland in a public campaign to fight against skin cancers.

Pupils' Perceptions: Differences between Gender, Age and School Level

We distinguish first perceptions of girls and boys. As the results (see Table 6) show, the perceptions of the boys are more positive than those of girls for all three subdimensions (RA: $d=0.51$; IE: $d=0.85$; SC: $d=0.50$; in all cases $p\leq 0.01$, except for SC: $p=0.05$); the differences are most pronounced for intrinsic interest. These results are consistent with international findings about gender preferences regarding the sciences (Zwick and Renn 2000; Bøe et al. 2011). Comparing units, *Greenhouse* is slightly more interesting for the boys, while the perception of girls does not show any noteworthy difference.

Two other differential characteristics of our sample have been investigated: differences in perceptions between the class grade of the pupils (8th grade versus 9th grade) and between the school level (high level (A) versus low level (B)).

Even if some values are numerically different, e.g. for RA in *Sunscreens* and *Greenhouses*, and for IE in *Clothes*, the differences do not attain statistical significance (at the $p=0.05$ level). Note that self-concept is higher for 8th grade pupils, which can be understood by the fact that these pupils have chosen the scientific option.

The same holds for the differences between the school levels: there are numerical differences, which are however not statistically significant, even if the results for total motivation for the average over the PISA units ($p=0.068$) and for *Clothes* ($p=0.055$) are close to it. Note that the overall tendency is that the perception of the B pupils is better than that of the A pupils, so the perception of the lower performing group is at least not worse than that of the better group.

Teachers' Perceptions

Within the teachers' sample, evaluation of total motivation is globally much higher than within the pupils' sample,

Table 7 Teachers perceptions (averages and standard deviations; values are percentages of maximum score, see 4.3) of the five released units linked with physics for the two subdimensions (RA, IE) and their sum (Mot)

Values in %	RA	IE	Mot
<i>Sunscreens</i>	74 (22)	74 (23)	74 (18)
<i>Greenhouse</i>	83 (18)	81 (21)	82 (17)
<i>Clothes</i>	58 (32)	47 (35)	53 (31)
PISA (3 units)	72 (20)	68 (20)	70 (18)
<i>Grand Canyon</i>	61 (26)	58 (27)	60 (22)
<i>Acid rain</i>	70 (25)	64 (33)	67 (27)
PISA (5 units)	69 (19)	65 (20)	67 (20)

ranging from 53% for *Clothes* to 82% for *Greenhouse* to compare with 47% for *Clothes* and 49% for *Greenhouse* (see Table 7). Moreover, we note that also for teachers like for pupils, *Greenhouse* has highest total motivation, but the differences among the units are more pronounced in teachers' perceptions.

The results show first that *Clothes* is lowest in all aspects for teachers. Even if its perceived authenticity reaches a value higher than among pupils (58 vs. 46%), its intrinsic interest for teaching (47%) is considered low, although its second item dealing with electricity is part of the curriculum. According to the teachers, *Greenhouse* and *Sunscreens* have a high teaching interest. An interpretation of these differences on the unit level is given in 6.1.

As this study is inquiring about possible differences in teachers' and pupils' perceptions of authenticity and intrinsic interest (T-P comparison), we have analysed significance levels and effect sizes for these dimensions for all three units assessed within both samples. Due to the simple variable structure, tests were carried out as t-tests, after checking for prerequisites (see 4.3). These differences are highly significant ($p < .001$ in all cases), and effect sizes generally very large (range $d \approx 1.3$ up to 1.9; see Table 8), with the IE differences being even more pronounced than the RA differences. Comparing units, the only exception is *Clothes*, where no significant differences were found.

Teachers' Assumptions about Pupils' Perceptions

We now turn to the second part of the teachers' questionnaire, investigating their assumptions about pupils' per-

Table 8 Effect sizes (Cohen d) for differences between pupils' perceptions and teachers' perceptions (P-T comparison). The significance level of all differences is $p < .001$, with the exception of *Clothes*, where no significant differences were found

	RA	IE	Mot
<i>Sunscreens</i>	1.30	1.94	1.52
<i>Greenhouse</i>	1.35	1.87	1.53
<i>Clothes</i>	–	–	–
PISA (3 units)	1.34	1.60	1.37

ceptions. Comparing units, the order of these assumptions closely parallels that of their own perceptions for both RA and IE (cf. 5.3), up to the interchange of *Grand Canyon* and *Clothes* for RA.

The comparison between teachers' assumptions about pupils' perceptions (TP) and the actual pupils' perceptions (P) (Table 9) was carried out as in Sect. 5.3 as t-tests (after checking for prerequisites, see 4.3). It again shows statistically significant differences, with noticeable effect sizes (RA: $d=0.66$, $p=0.01$; IE: $d=0.85$, $p=0.001$). As for the T-P comparison, IE differences are more pronounced than the RA differences. Comparing individual units, only *Clothes* did again not show significant differences, while the most pronounced differences occurred for *Greenhouse*.

Figure 1 summarizes the results of the three perspectives investigated here: pupils' perceptions (P), teachers' perceptions (T), and teachers' assumptions about pupils' perceptions (TP). It shows clearly the decreasing perceived authenticity, intrinsic interest and the combination of both, going from personal perceptions of teachers to personal perceptions of pupils, with the teachers' assumptions about pupils' perceptions in between.

Discussion

Comparison between Units

Within both the teachers' and the pupils' sample, the overall ordering of the units (all five for teachers, three for pupils) motivational potential is in line with the basic idea of CBSE (i.e. the stronger the contextualization, the better for motivation) and the specific features of these units with respect to the societal and curricular background in Geneva. On the one hand, this can explain the high score of *Greenhouse*, which is a vivid question in our society, as well as of *Sunscreens* because the danger of skin cancer by sunburns is treated every spring in Switzerland in the mass media. Moreover, both units are about the interpretation of experimental results in a way close to the spirit of the physics curriculum in Geneva. Note that the findings about teachers' assumptions about pupils' perceptions validate PISA's choice of the units *Sunscreens* and *Greenhouse* for the population of this study, as teachers choose them as the most authentic and interesting ones for pupils.

Grand Canyon with its conventional items on the effect of the freezing water on the rocks and on the explanation of the presence of fossil animals in the mountains is not a problem of current interest. The problem of the *Acid rain* has lost its importance in the media these last years. Finally, *Clothes* has the lowest score for both pupils and teachers, and the latter are little convinced about the authenticity and interest of the subject for pupils. But the unit has, at face value, some

Table 9 Pupils’ perceptions (P) and teachers’ assumptions about pupils’ perceptions (TP) differences (P-TP comparison; values are percentages relative to maximum score, see 4.3). The significance level of all differences is $p < 0.05$, with the exception of *Clothes* (no significant difference)

	RA				IE				Mot			
	TP	P	<i>p</i>	Cohen <i>d</i>	TP	P	<i>p</i>	Cohen <i>d</i>	TP	P	<i>p</i>	Cohen <i>d</i>
Sun.	59 (30)	47 (21)	0.04	0.61	60 (22)	37 (19)	<0.001	1.20	59 (23)	44 (19)	0.004	0.79
Green.	73 (16)	52 (23)	<0.001	0.89	65 (22)	39 (22)	<0.001	1.18	69 (17)	49 (22)	<0.001	0.89
Cl.	49 (32)	46 (24)	0.36	–	37 (30)	41 (20)	0.30	–	43 (29)	45 (23)	0.40	–
PISA (3 units)	60 (20)	48 (18)	0.01	0.66	54 (18)	39 (18)	0.001	0.85	57 (17)	46 (17)	0.007	0.62
Gr. C.	48 (20)				49 (25)				48 (18)			
A. rain	56 (26)				50 (30)				53 (26)			
PISA (5 units)	58 (17)				54 (19)				56 (17)			

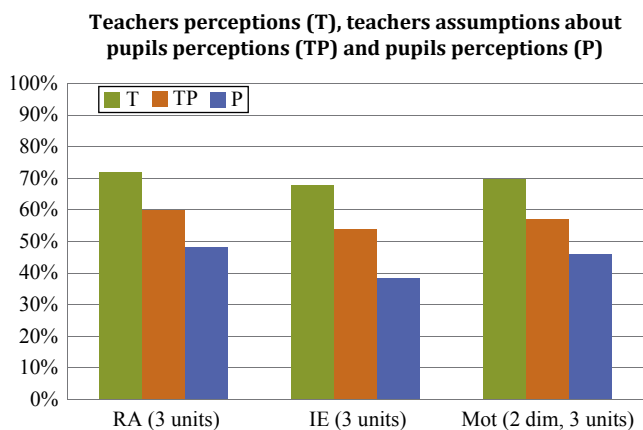


Fig. 1 Teachers’ perceptions, teachers’ assumptions about pupils’ perceptions and pupils’ perceptions for the three units, see Table 7 and 9: Pupils’ perceptions (P) and teachers’ assumptions about pupils’ perceptions (TP) differences (P-TP comparison; values are percentages relative to maximum score, see 4.3). The significance level of all differences is $p < 0.05$, with the exception of *Clothes* (no significant difference) and text

motivational potential: it is about an innovative and somehow surprising topic, which already nowadays belongs to a field where important technological progress is made, with many very useful applications to be expected in the future. It is obviously interesting to note, that such a potential can fail to be realized. Our interpretation of this low score is a mixture of two reasons. It has a rather long introducing text, two questions of the unit presenting relatively little interest (with regard to the long text), and it is also the last unit in the pupils’ questionnaire (with most blank answers). With the present data, it is not possible to analyse further the relative weight of these reasons on the level of the individual unit.

On the other hand, the fact the differences of the perceived motivation are less pronounced for pupils than for teachers lends itself to a plausible interpretation: the latter have all a strong disciplinary background in science (master up to several years of post-doc) and thus are well aware of the relative importance of the scientific issues treated in the different units.

Comparison of Teachers’ Perceptions, Pupils’ Perceptions, and Teachers’ Assumptions about Them

Our findings reveal first considerable differences between pupils’ and teachers’ perceptions. They are very large, as measured by Cohen *d* (across units, RA: $d = 1.34$, IE: $d = 1.60$, see Table 8) but this is not too surprising, as noted above. But the question is then, whether teachers sufficiently adjust these perceptions, when asked about pupils.

Second, the findings of this study show a still pronounced overestimation of pupil’s interest and perceived authenticity by physics teachers. These differences are again considerable (across units, RA: $d = 0.66$ IE: $d = 0.85$, see Table 9). All the surveyed teachers teach nowadays in secondary school (or formerly did for two of them). Therefore, this overestimation cannot be related to a mere ignorance of this age group of pupils. We interpret it as a sign that teachers, with their greater interest for and awareness of the importance of science, cannot, when thinking about pupils, totally quit their own present position. This overestimation can be a real problem in the everyday classroom work. Certainly, it is important that the teacher shows to the pupils his or her interest and enthusiasm towards the subject he or she teaches. But the danger to “lose contact” with the pupils by choices of situations that have little interest for them is big. So for the teacher there is a thin line between the aim of letting pupils discover new domains particularly in science and the necessity to remain close to their real, genuinely personal questions. As an example, some questionnaires (non valid) had comments like “this is bullshit!” Clearly, the PISA units did not meet the interest of these pupils.

Of course, young people are not naturally interested in and will not spontaneously discover many valuable topics—one important reason why they have to *learn*. Often in classrooms learner’s questions have to be elicited. On the basis of the present findings, one can furthermore state that (i) one has to ask the learners, as opposed to teachers (and researchers) whether they consider some topic or problem as really “authentic” (i.e. related to their lives), and that (ii) this feeling of authenticity is not guaranteed even for the PISA units, which were developed with such a great care and expertise.

Limitations of the Present Study

The present study has shown that pupils have a limited perception of the authenticity of the PISA units studied. We have discussed reasons inherent to the units themselves for this fact. Our findings, however, are subject to some limitations discussed now.

First, the pupils' survey had to be conducted in a last or the last but 1 week before the end of the school year for practical reasons (available test time in a quite tight teaching program). Therefore the pupils were perhaps not in the best mood towards school in general, and physics learning in particular, thinking already of the forthcoming holidays.

Second and more important, there was no attempt of teaching physics (or more generally, science) on the base of PISA like units before the test, neither for PISA itself, nor for the present investigation. Pupils had to pronounce themselves on the authenticity of the units in a sort of "stand alone" test situation, and it could be that their perceptions would change after having worked in a learning situation with this type of problems (actually, it could change in both directions). But then, this "mere-test situation" is a limitation to PISA as well, and the present study aimed at an investigation of perceived authenticity and interest in the PISA setting as such. It remains an interesting open question, how a real intervention based on PISA-like units would influence these factors (and learning, of course), but this is beyond the scope of the present study.

Conclusions

With regard to the research questions of this contribution, we found first that pupils perceived the reality connection/authenticity (RA) as well as intrinsic interest/engagement (IE) related to the PISA physical science units at best at a medium level, not clearly showing the positive perceptions which should be expected according to the PISA philosophy (see e.g. Fensham 2009). Second, teachers have significantly better perceptions of both RA and IE than pupils, with very large effect sizes ($d \geq 1.3$). Third, even though they correctly think that pupils' perception are lower than theirs, they still considerably overestimate them ($d \geq 0.6$). We want to underline this finding, as it empirically validates and quantifies a probably frequent illusion among science educators and teachers: they project their own interest on pupils and are therefore overestimating the real perceptions (and needs) of pupils. While this overestimation is probably a risk in general, our results apply to PISA units in particular. PISA places large value, and justly so, on authentic contexts, and the interest generated by these, yet pupils' perceptions seem not to be as positive as the PISA philosophy seems to suggest. The consequence is not that PISA was "bad" in this respect, but that authenticity cannot be imposed from outside (just as

interest), and that it has to be assessed whether a given topic, problem or learning activity is really perceived as such by the learners. An interesting open question is then whether pupils' limited perceptions of authenticity and interest of the PISA units would change if these were integrated in a real learning sequence or whether they are due to the topics chosen.

Most importantly, however, we feel that the above statement about the impossibility of imposing authenticity, and the necessity to assess it, has implications both for research and classroom practice: as context-based science education plays such a large role in current research, for PISA, and well beyond, one has to take care that a supposed "authenticity" of these contexts is not considered as independent variable. Rather, it is already a *dependent* variable of external factors, which actually can be independently varied (topics, format, etc.). Only once authenticity is established in this sense as such a dependent, intermediate variable, its influence on dependent outcome variables such as attitudes and learning can be studied.

In classroom practice, a similar statement holds: teachers must be aware that it is pupils' perceived authenticity which matters, and that there is a considerable overestimation bias concerning it, even among people well experienced with science teaching of a given age group.

Finally, both for researchers and teachers, a handy, validated instrument of perceived authenticity is useful, and we feel that the questionnaire used here (Vogt, Kuhn and Müller 2011) has passed its proof of concept. Further use and improvement in cooperation with colleagues both in research and classroom practice would be welcome.

On the bottom line, we come back to one of the classical challenges of (not only) science teaching. This challenge turned out not even to be self-evident in the PISA approach, for which the present study offered an empirical, quantitative validation as "added value", and which might be expressed, with respect to the *genus loci* of Geneva, as follows (E. Claparède, Geneva, 1873–1940; as cited by Aebli 1983): *Une leçon doit être une réponse*, i.e. *a lesson has to be an answer [for learners]*—and learning and testing *problems* have to be those of learners as well.

Acknowledgements Very valuable support for the data analysis by M. Germond (Genève) and R. Hettmansperger (Ludwigsburg) is gratefully acknowledged.

References

- Aebli, H. (1983). *Zwölf Grundformen des Lehrens*. Stuttgart: Klett-Cotta.
- Bennett, J., Lubben, F., & Hogarth S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education*, 91(3), 347–370.

- Bøe, M. V., Henriksen, E. K., Lyons, T., & Schreiner, C. (2011). Participation in science and technology: Young people's achievement-related choices in late modern societies. *Studies in Science Education*, 47 (1), 37–72.
- Cariou J.-Y. (2010). Tentative de détermination de l'authenticité des démarches d'investigation. In C. Loisy, J. Trgalova, & R. Monod-Ansaldi (Eds), *Actes des journées scientifiques DIES 2010: Démarches d'investigation pour l'enseignement des sciences: quelles ressources pour les enseignants? Quelle place pour le travail collectif?* Lyon: INRP. <http://www.inrp.fr/publications/edition-electronique/dies2010/DIES2010-actes-complet-17-12.pdf>. Accessed 11 Jan. 2014.
- CTGV (Cognition and Technology Group at Vanderbilt). (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher*, 19(6), 2–10.
- Fensham, P. J. (2009). Real world contexts in PISA science: Implications for context-based science education. *Journal of Research in Science Teaching*, 46, 884–896.
- Hattie, A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Herrington, A., & Herrington, J. (2006). *Authentic learning environments in higher education*. Hershey: Information Science Publishing.
- Hoffmann, L., Häußler, P., & Peters-Haft, S. (1997). *An den Interessen von Mädchen und Jungen orientierter Physikunterricht. Ergebnisse eines BLK-Modellversuches*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel.
- Jenkins, E. W. (2006). The student voice and school science education. *Studies in Science Education*, 42, 49–88.
- Kuhn, J. (2010). *Authentische Aufgaben im theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht*. Wiesbaden: Vieweg + Teubner.
- Kuhn, J., Müller, A., Müller, W., & Vogt, P. (2010). Kontextorientierter Physikunterricht. Konzeptionen, Theorien und Forschungsergebnisse zu Motivation und Lernen. *PdN-PhiS* 59, 5, 13–25.
- Mims, C. (2003). Authentic learning: A practical introduction and guide for implementation. *The Meridian Journal*, 6(1). http://www.ncsu.edu/meridian/win2003/authentic_learning. Accessed 24 Oct. 2014.
- Moreau, J. (2008). Essai d'interprétation des résultats en fonction du contexte de l'élève et de son attitude par rapport aux sciences. In C. Nidegger (Ed), *PISA 2006: Compétences des jeunes romands Résultats de la troisième enquête PISA auprès des élèves de 9e année* (pp. 149–170). Neuchâtel: IRDP.
- Müller (2009). 'Modified Anchored Instruction' im Naturwissenschaftlichen Unterricht: Ein Interventions- und Forschungsprogramm Contribution cluster. In D. Höttecke (Ed), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik. GDCP-Tagungsband 2009*. Münster: LIT.
- Muckenfuß, H. (1996). Orientierungswissen und Verfügungswissen. Zur Ablehnung des Physikunterrichts durch die Mädchen. *Unterricht Physik* 31, 7, 20–25
- Murphy, P., & Whitelegg, E. (2006). Girls in the physics classroom: A review of the research on the participation of girls in physics. Institute of Physics, London, UK. <http://oro.open.ac.uk/6499/>. Accessed 20 May 2013.
- OECD. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
- OECD. (2007). *PISA 2006. Science competencies for tomorrow's world, volume 1: Analysis*. Paris: OECD.
- Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of Europe*. Bruxelles: EU Directorate-General for Research, Science, Economy and Society.
- Rost, D. (Ed). (2006). *Handwörterbuch Pädagogische Psychologie*. Weinheim: Beltz.
- Shaffer, D. W., & Resnick, M. (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, 10, 195–215.
- Scott, T. F., Schumayer, D., & Gray A. (2012). Exploratory factor analysis of a force concept inventory data set. *Physical Review ST Physics Education Research*, 8, 020105(1–10)
- Vogt, P. (2010). *Werbeaufgaben im Physikunterricht: Motivations- und Lernwirksamkeit authentischer Texte*. Wiesbaden: Vieweg + Teubner Verlag.
- Vogt, P., Kuhn, J., & Müller, A. (2011). Kreuzvalidierung verschiedener Interessenstests – Messen etablierte Instrumente das Gleiche? Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie. In D. Höttecke (Hrsg.), *GDCP-Tagungsband 2010*. Münster: LIT.
- Weiss (2010). L'enseignement des sciences au secondaire obligatoire en Suisse romande, au regard des enquêtes internationales sur la culture scientifique des jeunes. *Revue Suisse des Sciences de l'Education*, 32/3, 393–420.
- Weiss (2015). (accepted for publication). PISA-sciences est-il IBL-compatible? To appear in *Recherches en didactique*, 19.
- Zahner Rossier, C., & Holzer, T. (2007). *PISA 2006. Les compétences en sciences et leur rôle dans la vie*. Rapport national. Neuchâtel: OFS/CDIP.
- Zwick, M., & Renn, O. (2000) *Die Attraktivität von technischen und ingenieurwissenschaftlichen Fächern bei der Studien- und Berufswahl junger Frauen und Männer*. Stuttgart: Akademie für Technikfolgenabschätzung.