INT. J. SCI. EDUC., 19 MARCH 2004, VOL. 26, NO. 4, 425–445

RESEARCH REPORT

How can we identify replies that accurately reflect students' knowledge? A methodological proposal

N. Marín, Depto. de Didáctica de la Matemáticay de las Ciencias Experimentales, Facultad de Ciencias de la Educación, Universidad de Almería; e-mail: nmarin@ual.es; E. Jiménez Gómez, Depto. de Didáctica de las Ciencias Experimentales, Facultad de Educación, Universidad de Murcia and A. Benarroch, Depto. de Didáctica de las Ciencias Experimentales, Escuela Universitaria de Melilla, Universidad de Granada, Spain.

The catalogue of conceptions that students are said to have concerning the different topics of the science curriculum is so great that some authors consider this line of research to be exhausted. However, others insist on the need to re-examine students' conceptions in order to better describe them using new theoretical, contexts and research methods. We present an integrated set of guidelines and methodological techniques for selecting those replies that best reflect students' knowledge.

Introduction

The 1980s were a fruitful period for investigative activity into students' conceptions concerning different scientific topics (Driver et al. 1989, Carmichael et al. 1990, Hierrezuelo and Montero 1991, Duschl 1994), which has improved science teaching (Driver and Oldham 1986).

The 1990s saw new approaches to research into students' conceptions, the consequence of which were improvements in the way information is gathered from students and new strategies for increasing the reliability and validity of the results obtained. Some of these changes involved the following:

- Methodological changes, basically substituting pencil and paper questionnaires by interviews or semi-structured dialogues with clear metacognitive intentions (Gutiérrez and Ogborn 1992, Vosniadou and Brewer 1992, Benarroch 1998). Such interviews are also used to look for a certain coherence and regularity in the students' replies when they are asked about a variety of physical facts.
- The use of questionnaires with wider age ranges representing different educational levels (for example, Galili and Bar 1992, Reynoso et al. 1993, Bar et al. 1994, Kuiper and Mondlane 1994, Montanero et al. 1995, Benson et al. 1993), thus obtaining a 'cross-age' view of students' conceptions.
- The use of techniques to evaluate, at least partially, the data obtained and thus appraise the reliability and validity of students' conceptions (see, for

International Journal of Science Education ISSN 0950–0963 print/ISSN 1464–5289 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/095069032000097352

example, Finegold and Gorsky 1991, Thijs 1992, Reynoso et al. 1993, Kuiper and Mondlane 1994, Montanero et al. 1995, Vosniadou and Brewer 1994).

As a continuation of this and in an attempt to improve research into students' conceptions, this contribution presents a series of guidelines and methodological techniques for obtaining valid and reliable information that better represents the knowledge of students concerning certain well-defined topics.

Preliminary questions

To give meaning to any received message (empirical data, perception, verbal message, etc.) and to respond to any question or problem situation, a student will use his/her previous knowledge. If no such knowledge exists, there are several possibilities. The response may then be:

- Induced or suggested by the question asked. For example, a student may well indicate the direction in which the current flows in an electrical circuit if asked a question like 'Which way does the electric current flow?'
- Dependent on what is perceived. It is common for a student to be carried along by the text of a question or by what he/she understands about a given situation to construct an answer. For example, a student asked to draw the vector force at different points of the trajectory of a coin tossed into the air may respond by drawing a series of arrows even if he/she has no notion of what force is (see Watts and Zylbersztajn 1981). Is it possible to say from such a reply that the student clearly distinguishes between force and velocity?
- Random or simply invented. For example, a student may have a certain notion about what force is, which will enable him/her to interpret sufficiently well a large number of situations, although this knowledge is insufficient to explain whether a torch weighs less when it is switched off than when it is switched on, or to explain the force acting on an object that is dropped on the moon's surface (see Watts and Zylberstajn 1981). A student may not know the meaning of the word 'ion' when asked to define it and may provide an invented answer (see De Posada 1993).

We think that replies such as these can be distinguished from those in which the student's knowledge intervenes, and this is the object of our contribution since not all replies will be reliable and valid for identifying previous knowledge.

To differentiate one type of reply from others, we must accept that the schemes of knowledge of an individual will have certain characteristics of stability and coherence (Piaget 1978, Pascual-Leone 1979, Case 1983, Marín 1994a). If this is so, it is reasonable to think that in a series of replies arising from such schemes, certain regularities will be identified. At the same time, if the replies seem regular it will be because these replies will have been generated by one or several knowledge schemes.

Finally, we must define the meaning of the terms conception and scheme, which will be used in this paper. The word scheme refers to a construct that forms part of the unobservable cognitive web of the subject, while conception refers to the replies that show a certain degree of regularity and are an observable manifestation of students' cognition (see, for example, Brumby 1979, Nussbaum and Sharoni-Dagan 1983, Terry and Jones 1986, Brown 1989).

It must be stressed that this work is fundamentally methodological, since it offers a series of guidelines and methodological techniques for selecting those replies that best reflect students' knowledge.

Methodological proposal

This section consists of five parts. The first is related with the selection of the topic, the second with the methodological techniques used, the third with the application of these techniques to a specific topic, the fourth with the way in which data are collected, and the fifth with the treatment of the data.

The content

To demonstrate the methodology we have chosen a theme from mechanics in which concepts related with mechanical equilibrium are dealt with. The reasons for this choice are as follows.

- Students have a large baggage of daily knowledge that they have built up spontaneously from their interaction with situations of equilibrium (walking, building tower with blocks, placing objects, etc.).
- The whole range of methodological techniques proposed in the following can be applied, bearing in mind that the aim of the study is to test the methodology, rather than to learn about students' conceptions per se.

The initial problem a researcher has to deal with when designing a questionnaire is where to start! The first step, already mentioned, is related to the teacher's or researcher's knowledge of the topic, which should be reflected is the systematic approach applied to the facts and events to be used in the questionnaire.

A second step, complementing the previous one, is to break up the topic into different parts, so that each part becomes a unit of study. Despite the absence of fixed rules, it is evident that certain criteria have been used more than others. Some of the most used are:

- Contents that can easily and logically be broken up into different parts. For instance, *the senses* can be divided into five parts or can be studied globally.
- Topics such as 'the composition of forces' can be broken up, depending on the level of complexity: aligned forces of the same magnitude and direction or different direction, and so on.

Once the teaching topic has been chosen, it is advisable to confront students with a diversity of facts and events that underlie the content to be investigated and to allow them to express their knowledge on the topic (Marín 1994b, Benarroch 1998).

Methodological strategies

Different types of strategies must be used in each task so that the students' cognitive schemes are more comprehensively involved:

(a) The strategy of 'confrontation', which basically comprises two stages:

- Prediction stage, where the student is asked to anticipate possible outcomes for a physical fact or event.
- Empirical verification and confrontation stage, where the researcher or student carries out an experiment, the results of which are confronted with the answers given in the previous stage and with those given in this stage.

(b) Contextual variation, which consists of presenting different physical situations with the same underlying research topic. This basically comprises two stages:

- Relevant variations, which are modifications of some factors involved in the situation and which cause some kind of relevant alteration. Usually, but not necessarily, a relevant variation involves some change of form.
- Non-relevant variations, which are modifications of certain factors involved in the situation, which do not involve any kind of relevant alteration.

Similar techniques to these were used by Piaget (1946, 1975).

To design a questionnaire for a given research topic, the methodological strategies are applied in the following order:

- First, contextual variations are applied; that is, different situations that reflect the same underlying content. Each of these situations will give rise to a different task.
- Second, each task is begun with a prediction stage, where the student is asked to make different predictions as the different intervening factors change (relevant and non-relevant variations).
- Third, the student is shown the result of an experiment or is given information in the form of empirical data or drawing, which are compared with the initial replies of the prediction stage. This step continues with the relevant and non-relevant variation strategies.

The combination of the different strategies mentioned increases the possibility that students will apply their knowledge schemes more thoroughly when answering questions.

The degree of regularity in the students' replies can be evaluated using three criteria:

- Repetition, or the replies that remain unaltered despite the modifications of the physical situations introduced using both non-relevant and relevant variation strategies.
- Generalization, as observed from analogous replies to the physical situations principally constructed from contextual variation strategies.
- Adaptation of the replies to the factors intervening in the task (contextual variation and confrontation strategies).

Variation and confrontation strategies allow more information to be gathered from students in such a way that the answers show signs of repetition, generalization and adaptation. Note that both strategies used in the design and construction of the questionnaire that is used as protocol in the interview are based on the underlying model of cognitive organization, where the organization unit is a knowledge scheme

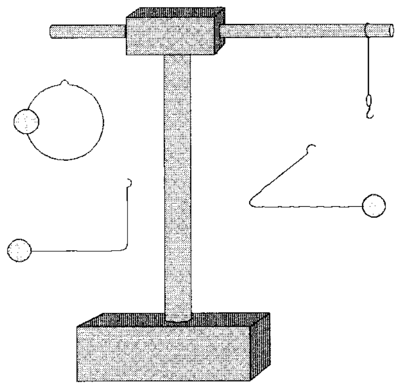


Figure 1. Objects to be hung from a hook.

with characteristics of stability and coherence (Piaget 1978, Pascual-Leone 1979, Case 1983, Marín 1994a).

Students' replies can be described in a student sample covering a broad age range to obtain representation of the different level of educational and mental development.

Applying the methodological techniques to a given topic

Aware of the difficulties implicit in our methodology, we illustrate the different methodological actions and orientations using our research into the evolution of students' knowledge schemes to explain mechanical equilibrium (Marín 1994b).

Each questionnaire is initially based on the systematic study of the topic. In Appendix 1 we present: task 1, which treats 'equilibrium when objects are allowed to hang freely'; task 2, the 'equilibrium of objects on a supporting structure' and task 3, a 'biweight in equilibrium'.

Examples of 'confrontation' strategy

- Task 1: Equilibrium of hanging objects
  - (i) Prediction stage: the replies to this task provide the variable VER_A ('VER' refers to the idea of verticality implicit in the force of gravity while the letter 'A' refers to replies given before (ante) hanging the shape from the hook). See task 1 (Appendix 1): questions 1a, 1b, 1c, 1d and 1e.
    The subject is asked to draw how three objects hanging from a hook will come to rest. The objects are composed of aluminium wire of differing shapes with a ball of plasticine (see figure 1 or Appendix 1). Note that modifications are made following the variation strategies already mentioned.
  - (ii) Empirical verification and confrontation stage: the replies to this task provide the variable VER_P (The letter 'P' refers to replies given after (post) empirical observation). See task 1 (Appendix 1): questions 1f, 1g, 1h, 1i, 1j, 1k and 1l.

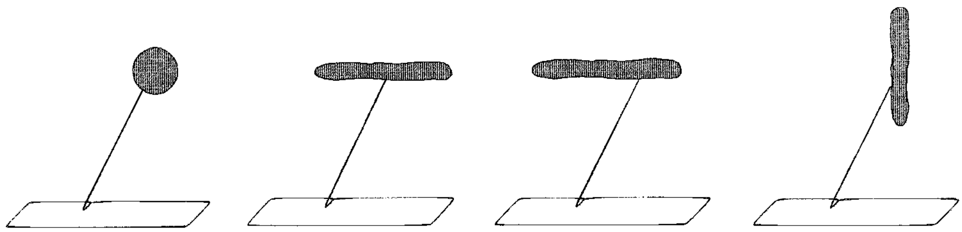


Figure 2. Shapes and positions of the plasticine on the support structure.

After the student has made a drawing of the possible positions of the shapes, these are suspended and the student is asked for the relevant explanations. The student is also asked to compare the explanation given in the prediction stage with those of empirical observations.

- Task 2: Equilibrium of objects on a supporting structure
  - (i) Prediction stage: the replies to this task provide variable MIC_A (MIC refers to the microphone shape of the structure). See questions 2a, 2b, 2c and 2d of Appendix 1, task 2. The same materials as in task 1 (aluminium wire and plasticine) are used to construct a wire support with a piece of plasticine attached to one end (see figure 2).
    The students are asked to predict the positions of equilibrium of the plasticine in relation with the support and the positions in which it would fall.
  - (ii) Empirical verification and confrontation stage: the replies to this task provide the variable MIC_P. See task 2 (Appendix 1): questions 2a, 2b, 2c and 2d. The student is asked what happens empirically when the plasticine and structure are in a certain position. In each position, after the confrontation stage, the student is asked for clarifications and causal explanations.

It was seen that the subjects suffered greater cognitive conflict when the plasticine (in elongated shape) was hung vertically with the centre of mass not outside the edge of the support (drawing on the right of figure 2).

- Task 3: Equilibrium of biweight position
  - (i) Prediction stage: the replies of this task provide variables BIW_A. See questions 3a, 3d, 3f and 3h of task 3 (Appendix 1). The object (designated biweight) is show in several positions (see figure 3) and the student is asked to predict in which it would remain in equilibrium if supported on the central wire.
  - (ii) Empirical verification and confrontation stage: the replies of this task provide variables BIW_P. See questions 3b, 3c, 3e, 3g and 3i. The ‘biweight’ (see figure 3) is first placed in the predicted position and the result is observed. The student is asked to compare the explanations given in the prediction stage with those given during the observation. The factors intervening in the situation are then modified to create

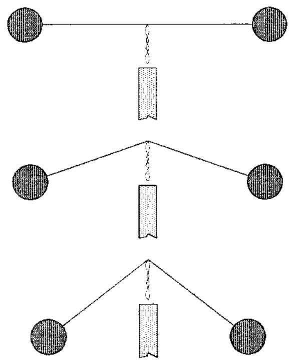


Figure 3. Some of the biweight positions.

cognitive conflict. For example, if the student thinks that a heavier weight would result in greater equilibrium, the investigator increases the weight, demonstrating that that greater equilibrium is achieved. The ball can also be removed, which is when the object falls.

Examples of variation strategies

In order to obtain, on the one hand, the widest possible information from the student concerning the subject under investigation and, on other hand, to be able to discern random or invented answers from those that involve (to whatever extent) his/her knowledge schemes, we have used what we shall call '*variation strategies*'.

In the protocol about 'situations of mechanical equilibrium', we present equilibrium in different contexts. Figures 1–3 show examples of physical situations that, although modified, reflect the same underlying content (the concept of verticality, equilibrium and centre of mass come into play).

- Relevant variations, which are modifications of some factors involved in the situation, which cause some kind of relevant alteration. Usually, but not necessarily, a relevant variation involves some change of form. For example, if we start from the biweight in equilibrium (see figure 3), the attaching wires can be bent slightly so that the centre of mass is slightly above its point of support, so that the biweight is now unbalanced.
- Non-relevant variations, which are modifications of certain factors involved in the situation, do not imply any kind of relevant alteration. For example, in the first task (see Appendix 1 or figure 1), several objects are made out of aluminium wire with a plasticine ball at the end. Because of the malleability of the wire, the objects can be shaped (see drawing), but in all cases, since the wire mass is insignificant compared with that of modelling plasticine, the ball always remains in the same vertical line as the hook. Therefore, variations in the way the wire is moulded are irrelevant in relation to the object's position when hung.

Obtaining data

By means of personal interviews, the protocol that is designed using *confrontation* and *variation* strategies ensures that each student provides a great amount of information. Videotape is advisable in order to extract all the necessary data for the research. We usually select 40 students for our studies with a broad age range (in this study 6–24 years) in an attempt to obtain a representations of the different replies of the students. This number (40) is justified by the fact that above 30 the answers given are repetitions of those previously registered and do not provide new information. However, the further 10 students interviewed provide a certain degree of security that no new empirical categories or modifications appear (see the fragment of interview in table 1, in which a student is looking at task 1 of Appendix 1).

Data treatment

Categorization and ordering (qualitative phase of data treatment) of students' answers. In order to classify the students' answers, two complementary and interdependent procedures have been used:

- The inductive procedure. The idea is to obtain a group of categories that are adapted to and faithfully include all the replies given by student during the interview, which Taber (2000) refers to as 'theoretical sensitivity'. The principal criterion lies in grouping the answers according to analogies and differences.
- The deductive procedure. The use of theoretical positions emphasizes some data more than some other and also gives order to the categories. For instance, the students' answers are analysed in order to look for new answers, distortion and, generally speaking, any kind of transformation of the observable aspects of the task.

At the same time that answers are being categorized, they are incorporated into a system of categories (empirical categories); this way, each individual's answer to a question or to a set of questions is given a category number, which represents the position of that answer in the system of hierarchical categories (table 2 shows the different categories obtained for variable VER_P; analogous tables were obtained for all the variables).

The first interview permitted us to construct the first system of categories, which helped us look for new data in subsequent interviews and with greater precision. The new data in turn served to differentiate, tune and even restructure the first system of categories. This cyclical approach continued until any new data provided by interviews could easily be assimilated into the system of categories without any need to alter them (Marín 1994b). This process of categorization is similar to that recently published by Taber (2000), although this author starts from different assumptions.

Quantitative data treatment. As previously mentioned, not all the students' answers are built from their knowledge schemes, which is why we need to evaluate the information obtained after it has been collected. Contrasting students' answers by variation (context, relevant or irrelevant) and confrontation allows us to determine the regularity of the answers and, therefore, to identify and describe their knowledge.

These variation strategies behave as control of the data collecting process, making it possible to differentiate between the information that comes from the students' knowledge schemes and answers given at random, invented, and so on.

Table 1. Interview with CIN (15 years), task 1.

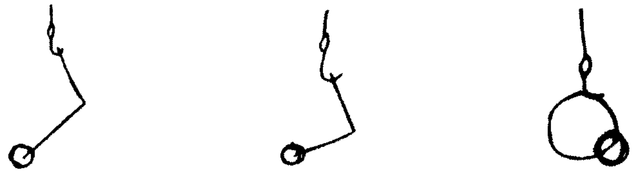
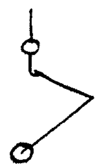
	CIN: pseudonym of student interviewed; P: questions and notes; R: interviewee's replies
P	Draw the position of the 'L' shape if we hang it from the hook. *[Drawing of the 'L' sloping downwards but far from the vertical].*
R	Because the ball is heavy it goes downwards when you hang it.  "L" Coathanger hoop *The student then draws the 'coathanger' less inclined than the 'L' and gives the same reasons. Finally the hoop is only inclined slightly (from the vertical). Both the hoop and the coathanger are inclined about 25°, while the 'L' reaches 50°. The different extent to which the objects is explained by the weight pulling down.*
P	*After the interviewee has drawn the predicted position, the interviewer hangs the 'L'.* Is that how you drew it?
R	Yes, exactly the same. *Since there are clear differences the interviewee is asked to make a new drawing. This time the ball is lower but not in the vertical.* 
P	Why is it like that?
R	Because the ball weighs so much and is attached, the weight takes it down.
P	But does the ball go to any special position?
R	In the same place as the hook but lower. *This new idea seems to have arisen from observation since it was not evident before. The 'L' is hung up again and the interviewer notices the differences. She draws it again, this time taking the vertical as reference.* 
P	Where's the ball on the 'coathanger' (shape of the wire)?
R	In the same place it's hanging from. *The hoop is hung on the hook and the interviewee is asked to draw it again if she has not predicted the position well.*
P	What's the connection between the hook and the ball? The interviewer draws a vertical line from the hook to the ball (look at this drawing) 
R	Because it's heavy, it's lower. *The interviewer draws the hoop again lower but still not vertical.*
P	Is the ball on the hoop in the same place as in the other objects?
R	No, . . [*Looks at it carefully*], yes.
P	*The three objects are hung at the same time, each on a different hook. The interviewee is told to look at the positions and is asked*: Have they got anything in common?
R	I can't see anything . . . it looks like the ball is under the hook on the 'L' and the coathanger, except the hoop where the vertical doesn't pass through the centre.

Table 2. Obtaining empirical categories for students' conceptions related with the variable VER_P.

Differentiating qualitative features	*Empirical category*	*Students*
Despite the obvious differences, the student is not aware of them and draws the ball somewhere in a wide area below the hook with no reference to the vertical. No mention of weight.	1	Cis, Van, Her, Lid, Lil, Lor, Pek
The student vaguely observes the differences and draws the ball below the hook but with no reference to the vertical. Weight is mentioned for the first time.	2	Ner, Nic, Ram
The student can appreciate the differences (with difficulty) and draws the ball below the hook with no reference to the vertical. He/she continues to mention weight.	3	Pin, Nor, Kin, Ton, Rez, Bed, Len
The subject can appreciate the differences (with difficulty). Observation leads him/her to notice the verticality involved (sometimes but not systematically).	4	Fan, Cin, Rom, Ser, Pat
Although the student sees the differences without difficulty, verticality is only appreciated sporadically (after observation).	5	Zip, Des, Gen, Fel, Car
The student notices the differences clearly and discovers the verticality on first observation and applies it thereafter.	6	Lis, Fin, Per, Rev
The student appreciates the difference (usually accurately) and is aware that shape is not relevant. Verticality is applied systematically although with slight problems for the position of the hoop.	7	Luk, Gal, Mer, Ter, Gon, Mif
The subject does not need to make the drawing since the ball vertical line immediately relates the ball and the hook.	8	Man, Sim, Dal

Data provided by *confrontation* and *variation* strategies, combined with statistical modules (principal components analysis, different cluster analyses and, especially, *correspondence analysis* (CA)), permit the classification of students' answers into relevant and less relevant, according to whether they reflect their knowledge to a greater or lesser extent. However, no matter how good statistical techniques are, if they are used without a protocol designed according to the confrontation and variation strategies, they will be less able to distinguish between relevant and non-relevant answers.

The statistical treatment of the variables based on empirical categories (table 2), makes it possible to obtain others, more precise, which we shall call 'structural categories' (table 3).

Of all the statistical techniques used (principal components analysis, different cluster analyses and, especially, CA), the CA provides most information and best reflects the structure of the data, since in one two-dimensional plane it shows the relations between the different categories of each of the variables in relative distances (Dixon et al. 1990).

Table 3. Description of the structural categories of VER_P.

Structural category	*Description Structural category*	*Empirical category*	*Students*
1	The differences are not always appreciated despite being quite pronounced. If differences are observed they are not seen accurately. The ball is drawn in a non specified area below the hook.	1	Cis, Van, Her, Lid, Lil, Lor, Pek
		2	Ner, Nic, Ram
2	The differences are seen with difficulty. The ball is drawn in a non specified area below the hook.	3	Pin, Nor, Kin, Ton, Rez, Bed, Len
3	The student has no or little difficulty in seeing the differences. After observation the vertical is noted but sporadically and not systematically.	4	Fan, Cin, Rom, Ser, Pat
		5	Zip, Des, Gen, Fel, Car
4	The differences are clearly seen. Verticality is noticed from the first observation and applied systematically.	6	Lis, Fin, Per, Rev
5	The differences are noticed, in some cases accurately, even before observation the vertical relationship between ball and hook is established. This is applied systematically except for some slight problems with the hoop.	7	Luk, Gal, Mer, Ter, Gon, Mif
		8	Man, Sim, Dal

Figure 4 shows the categories of the six variables, which are abbreviated for reasons of space (VER_A, VER_P, MIC_A, MIC_P, BIW_A and BIW_P). Figure 4 is the starting point for obtaining what we have termed structural categorization (table 3), which regroups the empirical categories (table 2) as a function of their distance from other categories. Intervals in which the different categories can be grouped were established on the two axes of the CA using a double criterion:

- To use the minimum number of intervals (otherwise the grouping effect would be lost).
- To obtain the maximum number of empty areas and the minimum number that contain the highest possible number of categories. In this way, the areas are closely representative of the data.

The *correspondence analysis*, along with the principal components and cluster analysis, also permitted effective and detailed characterization of the six variables corresponding to the three tasks by situating the qualitatively defined categories in five intervals of the principal axis. Table 4 presents the most relevant characteristics of the variables.

The nomenclature and meaning of the symbols used in table 4 are as follows:

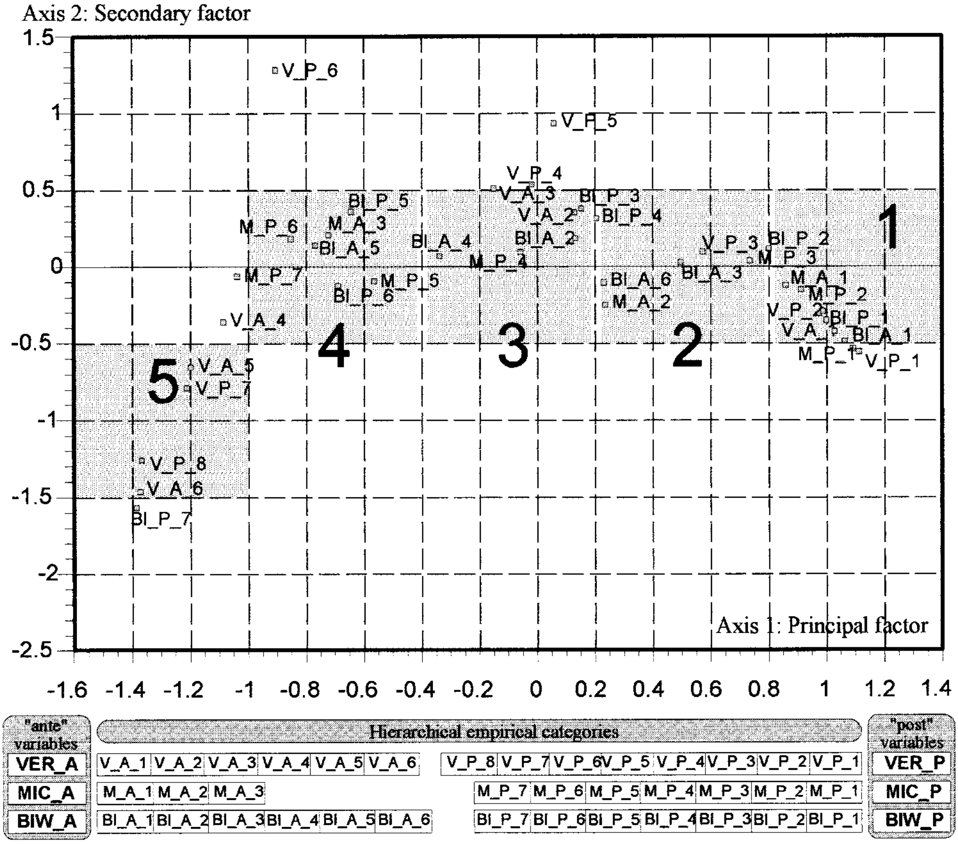


Figure 4. Categories of the six variables projected (correspondence analysis).

- The names of the variables and the tasks giving rise to them occupy the first and second columns, respectively.
- The third column shows the distribution of subjects according to the structural categories (1, 2, 3, 4 and 5, structural categories that are described in table 3 for the variable VER_P), and, below, of the third column show the original categories ordered according to their difficulty or complexity and deduced empirically from the replies of the students (see table 2). For example, for the variable VER_P, 10 students are classified in the structural category 1, these 10 students being classified in empirical categories 1 and 2 described in table 2.
- The first datum of the last column (DISCR.) expresses the intensity of the discrimination (+ indicating little and +++ a substantial degree of discrimination).
- The second datum (PARAL.) of the last column (parallelism to axis 1: principal factor) indicates the degree to which the original order of categories is maintained (+++ signifies a high degree of parallelism and – an inverted order). For example, for the variable BIW_A the negative sign (–) is due to the fact that the empirical categories 6 (table 2) have moved

Table 4. Description and characteristics of the variables on mechanical equilibrium.

Variable		*Tasks producing the variables*	*Distribution* 1	2	3	4	5	*Discr. Para*
VER_A	TASK 1	Verbal and graphical predictions concerning the position that several objects will adopt when hung from a hook	10	0	22	0	8	+ +
			1		2 3		4 6	+ +
VER_P		Verbal and graphical reactions once the actual position is compared to the prediction (confrontation stage)	10	7	10	4	9	+ + +
			1 2	3	4 5	6	7 8	+ + +
MIC_A	TASK 2	Prediction concerning the position of the plasticine on the support structure	14	7	0	19	0	+
			1	2		3		+
MIC_P		Empirical verification and confrontation results of which are confronted with the answers given in prediction stage	7	9	8	13	3	+ +
			1 2	3	4	5 6	7	+ +
BIW_A	TASK 3	Prediction concerning equilibrium of biweight position	2	20	8	10	0	–
			1	3 6	2 4	5		– –
BIW_P		Replies of empirical verification and confrontation stage of biweight position	7	4	12	15	2	+ +
			1	2	3 4	5 6	7	+ +

to structural category 3, meaning these has been an inversion of empirical categories.

From the characteristics of the six variables, the following can be appreciated:

- VER_A (variable referring to the students' replies for task 1 in the prediction stage) has categories in intervals 1, 3 and 5 (see figure 4), for which reason it discriminates the sample well without being among the best. The order of categories is maintained when projected on the CA space although two pairs are very close (see pairs [V_A_2, V_A_3] and [V_A_4, V_A_5] in figure 4). Although this is not a good variable, it offers substantial information on the conceptions held by students on verticality.
- VER_P (variable referring to replies to task 1 in the empirical verification and confrontation stage) is one of the best variables for our purposes. It is represented in the five intervals of figure 4, it discriminates very well, its original order is maintained in the CA projection and the content of its categories provide much information.
- MIC_A (prediction variable for task 2) is of little value. It is present in intervals 1, 2 and 4 of figure 4 (with a distribution of students of 14–7–0–19–0; see table 4) and therefore is a poor discriminator. Although

the order of its categories runs parallel to the principal axis of the CA, it has no components in third and last intervals. The information it provides is poor in comparison with that provided by VER_P.

- MIC_P (empirical verification and confrontation variable of task 2) is present in all the intervals of figure 4 whit a distribution of student of 7–9–8–13–3 (see table 4). It discriminates well. The information it provides is very significant.
- BIW_A (prediction variable of task 3) is defective in all possible ways. Although present in the first four intervals, its categories are strongly inverted compared to the original order (see empirical categories; table 2). The highest [BI_A_6] is situated in interval 2 near [BI_A_3]; the categories found in third interval are [BI_A_2] and [BI_A_4]; [BI_A_5] is in the fourth interval and [BI_A_1] in the first. Consequently it erroneously discriminates the subjects and it would be wrong to identify 'conceptions' with the information it provides.
- BIW_P (empirical verification and confrontation variable of task 3) has similar characteristics to MIC_P. It is present in all the intervals 7–4–12–15–2, it discriminates well and the CA projection does not break up its original order. The information it contains offers good information on the conceptions students have concerning the concept of 'centre of mass'.

The principal components analysis points to high correlation between most of the variables (one factor alone explains 60% of the variance with an internal consistency of $\theta = 0.9734$), which is confirmed by the position of the categories in the graph of the CA. There is, then, a common factor in the variables, suggesting that the subjects possess one or several knowledge schemes to explain the situations of mechanical equilibrium and which show through the changes of context and variation strategies.

Some variables correlate better than others with this common factor; in other words, some variables of some tasks reflect the students' knowledge schemes better than others. For example, such schemes are clearly manifested by VER_P, while BIW_A, which, among others, distributes the subjects in such a way that its suitability is questionable (e.g. good students show poor results).

When pairs of variables (before and after confrontation) are compared in detail, the aforementioned reasoning can be better understood.

- The correlation between VER_A and VER_P is 0.78. However, the second provides greater detail of the general knowledge students have concerning situations of static equilibrium. For example, category three of VER_A contains 22 subjects who consider that the weight causes the ball to fall slightly or a lot with respect to its original position to a not clearly identified place below the hook. The shape of each object means it does not fall much. The use of different objects, the comparison of the position in which the objects of different shape remain after hanging, and so on, lead the students to offer certain nuances that permit us to differentiate between those students who continue with their original opinions, those who, after observation, can appreciate verticality sporadically (but not systematically) and those who appreciate verticality immediately after observation and apply their knowledge systematically (see distribution of subjects according to variable VER_A and VER_P in table 4).

- The correlation between MIC_A and MIC_P is also high (0.85), and the same occurs as for VER_A and VER_P. MIC_A discriminates worse than MIC_P. For example, among those subjects who make bad predictions [MIC_A, 1], we can distinguish those who continuously fix on irrelevant factors [MIC_P, 1] from those who impose some sort of order on the data by delimiting an imprecise region on the upper part of the edge [MIC_P, 2]. For the group that alternates between correct and erroneous predictions [MIC_A, 2], those whose correct predictions are the result of intuition or chance [MIC_P, 1–2] can be distinguished from those who make predictions by means of a scheme, although sporadically [MIC_P, 3].
- From the correlation of 0.40 shown by BIW_A and BIW_P and given the 'qualities' of the first, we conclude that BIW_P contains information that badly reflects the students' true cognitive status.

Note that the confrontation and variation strategies of the qualitative stage, combined with the 'semi-quantitative' treatment of the ordinal category variable, make it possible to discriminate the information received. Qualitative techniques without their quantitative counterparts (or vice versa) would much lessen this discriminatory capacity.

Conclusions

(i) The variables obtained from the categorization of prediction replies (VER_A, MIC_A and BIW_A) discriminate less than variables obtained from the categorization when empirical verification and confrontation stage are used (VER_P, MIC_P and BIW_P).
To a certain extent, these findings are to be expected since the empirical verification and confrontation situations provide more time and opportunity to think than the prediction situations. In other words, the empirical verification and confrontation replies contain more information or reflect better students' knowledge.

(ii) Some variables, such as BIW_A, do not discriminate the students' replies very well, and some students who in other variables were included in the highest are included in the lowest categories in this variable. This means that the prediction replies given in task 3 should be ignored when it comes to identifying and describing the conceptions of students.

(iii) It must also be admitted that the variables constructed on some of the presented replies, for example VER_A, offer good results but cannot be compared with the variables based on the confrontation replies, VER_P. The information obtained from the empirical verification and confrontation replies is more likely to reflect students' knowledge than the information obtained from the prediction replies.

(iv) If we accept that information provided by the pencil and paper questionnaires is in most cases similar to that provided by the prediction replies in the personal interviews, we can affirm that, generally speaking, the pencil and paper questionnaire with its marked academic influence (very similar in this respect to academic examinations) offers a more dispersed and less coherent image of students' knowledge (see Taber 2000) than that offered by empirical verification and confrontation replies (in personal interviews), which activate the knowledge schemes of the students to their full extent.

(v) The findings show that students are better classified for knowledge level when given the opportunity to express their knowledge through the confrontation replies.

(vi) Although we have shown that not all the replies given by students have the same importance, it is necessary to analyse all the data provided by the students. However, student conceptions can only be deduced from the data shown to be the most reliable. In this regard the following should be borne in mind:

- Variation (contextual, non-relevant and relevant) and confrontation strategies, when used as a protocol for a personal interviews, provide information that permits the replies based on knowledge schemes to be separated from those that are invented or given randomly and that have little or nothing to do with such schemes.
- Information must be gathered from subjects of different ages, so that the cross-age of students' conceptions can be determined. In this way, the different stages through which students' knowledge passes to reach a given concept may be determined.
- The best way of obtaining data is to use a protocol designed and constructed using variation and confrontation strategies in a personal interview.
- The data provided by the confrontation and variation strategies, combined with statistical modules such as principal components analysis, different cluster analysis and, especially, CA, make it possible to separate the most relevant data from that which is not, as has been seen in the presented study.

This methodology described cannot be applied in the classroom because of its thoroughness, complexity and time-consuming nature, but it does permit multiple choice questionnaires to be drawn up, in which each question is constructed from the tasks similar to those of Appendix 1 and each option is based on the categorized answers given by the students in the interview. This type of test accurately represents students' knowledge and can easily be computerized and evaluated (Cano Sánchez et al. 1998). Furthermore, the result can readily be applied by the classroom teacher.

Acknowledgement

The authors are grateful to the DGICYT for funding this study (project BSO2001–0496).

References

Bar, V., Zinn, B., Goldmuntz, R. and Sneider, C. (1994). Children's concepts about weight and free fall. *Science Education*, 78(2), 149–169.

Benarroch, A. (1998). Las explicaciones de los estudiantes sobre las manifestaciones corpusculares de la materia. Tesis inédita (Facultad de Educación, Universidad de Granada).

BENSON, D. L., WITTROCK, M. C. and BAUR, M. (1993). Students' preconceptions of the nature of gases. *Journal of Research in Science Teaching*, 30(6), 587–597.

BROWN, D. E. (1989). Students' concept of force: the importance of understanding Newton's third law. *Physics Education*, 24, 353–358.

BRUMBY, M. (1979). Problems in learning the concept of natural selection. *Journal of Biological Education*, 13(2), 119–122.

CANO SÁNCHEZ, J., JIMÉNEZ GÓMEZ, E., MARÍN, N. and SOLANO, I. (1998). Identificación y descripción de las explicaciones de los alumno mediante una prueba de opción múltiple obtenida a partir de los resultados de entrevistas. In C. Martínez and S. García (eds.) *Actas de los XVIII Encuentros de DCE* (La Coruña: Servicio de Publicaciones de la Universidad de la Coruña), 261–272.

CARMICHAEL, P., DRIVER, R., HOLDING, B., PHILLIPS, I., TWIGGER, D. and WATTS, M. (1990). *Research on Students' Conceptions in Science: a Bibliography* (Leeds: University of Leeds, Children's Learning in Science).

CASE, R. (1983). *El desarrollo intelectual: una reinterpretación sistemática*. In M. Carretero and J. A. Madruga (eds.) *Lecturas de psicología del pensamiento. Razonamiento, solución de problemas y desarrollo cognitivo* (Madrid: Alianza Editorial), 339–362.

DE POSADA, J. M. (1993). Concepciones de los alumnos de 15–18 años sobre la estructura interna de la materia en el estado sólido. *Enseñanza de las Ciencias*, 11(1), 12–19.

DIXON, W. J., BROWN, M. B., ENGELMAN, L. and JENNRICH, R. I. (1990). *BMDP Statistical Software Manual*, Vols. I and II (Berkeley, CA: University of California Press).

DRIVER, R. and OLDHAM, V. (1986). A constructivist approach to curriculum development in science. *Studies in Science Education*, 13, 105–122.

DRIVER, R., GUESNE, E. and TIBERGHIEN, A. (1989). *Ideas científicas en la infancia y la adolescencia* (Madrid: Morata/M.E.C.) [original version (1985). *Children's Ideas in Science* (London: Open University Press)].

DUSCHL, R. A. (1994). Editorial policy statement and introduction. *Science Education*, 78(3), 203–208.

FINEGOLD, M. and GORSKY, P. (1991). Students' concepts of force as applied to related physical systems: a search for consistency. *International Journal of Science Education*, 13(1), 97–113.

GALILI, I. and BAR, V. (1992). Motion implies force: where to expect vestiges of the misconception? *International Journal of Science Education*, 14(1), 63–81.

GUTIÉRREZ, R. and OGBORN, J. (1992). A causal framework for analysing alternative conceptions. *International Journal of Science Education*, 14(2), 201–220.

HIERREZUELO, J. and MONTERO, A. (1991). *La ciencia de los alumnos, 'Su utilización en la didáctica de la Física y Química'* (Vélez Málaga: Elzevir).

KUIPER, J. and MONDLANE, E. (1994). Student ideas of science concepts: alternative frameworks? *International Journal Science Education*, 16(3), 279–292.

MARÍN, N. (1994a). Elementos cognoscitivos dependientes del contenido. *Revista interuniversitaria de formación del profesorado*, 20, 195–208.

MARÍN, N. (1994b). Evolución de los esquemas explicativo en situaciones de equilibrio mecánico. Tesis inédita (Granada: Facultad de Educación, Universidad de Granada).

MONTANERO, M., PÉREZ, A. L. and SUERO, M. I. (1995). A survey of students' understanding of colliding bodies. *Physics Education*, 30(5), 277–283.

NUSSBAUM, J. and SHARONI-DAGAN, N. (1983). Changes in second grade children's preconceptions about the Earth as a cosmic body resulting from a short series of audio-tutorial lessons. *Science Education*, 67(1), 99–114.

PASCUAL-LEONE, J. (1979). La teoría de los operadores constructivos. In Juan Delval (ed.) *Lecturas de psicología del niño* (Madrid: Alianza Universitaria), 208–228.

PIAGET, J. (1946). *Les notions de mouvement et de vitesse chez l'enfant* (Paris: P.U.F.).

PIAGET, J. (1975). *La composición de la fuerzas y el problema de los vectores* (Madrid: Morata) [original version [1973] *La composition des forces et le problème des vecteurs* (Paris: P.U.F.)].

PIAGET, J. (1978). *La equilibración de las estructuras cognitivas, 'Problema central del desarrollo'* (Madrid: Siglo XXI).

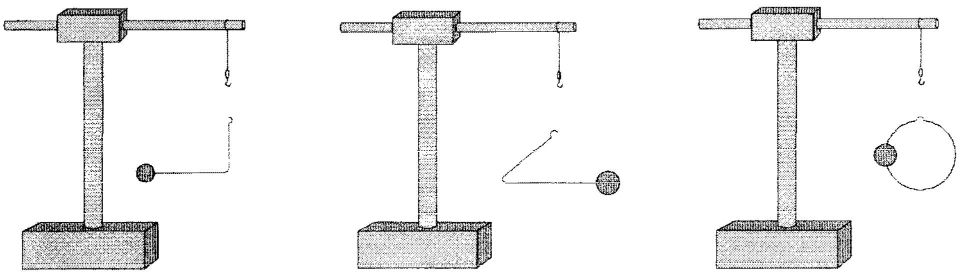
REYNOSO, E., FIERRO, E., GERDO TORRES, O., VICENTINI-MISSONI, M. and PÉREZ DE CELIS, J. (1993). The alternative frameworks presented by Mexican students and teachers concerning the free fall of bodies. *International Journal Science Education*, 15(2), 127–138.

TABER, K. S. (2000). Case studies and generalizability: grounded theory and research in science education. *International Journal of Science Education*, 22(5), 469–487.
TERRY, C. and JONES, G. (1986). Alternative frameworks: Newton's third law and conceptual change. *European Journal of Science Education*, 8(3), 291–298.
THIJS, G. D. (1992). Evaluation of an introductory course on 'force' considering students' preconceptions. *Science Education*, 76(2), 155–174.
VOSNIADOU, S. and BREWER, W. F. (1992). Mental models of the Earth: a study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585.
VOSNIADOU, S. and BREWER, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, 18, 123–183.
WATTS, D. M. and ZYLBERSZTAJN, A. (1981). A survey of some children's ideas about force. *Physics Education*, 16, 360–365.

Appendix 1: protocol about situations of mechanical equilibrium

Task 1: equilibrium of hanging objects with well-defined centres of mass (centre of plasticine ball)

Imagine what will happen if we suspend the plasticine ball on a wire with the followings shapes:



The L-shaped wire **The coathanger** **The hoop**

There is no need to draw the support, only the thread hanging from the support and the hanging object.

1a. Draw how the L-shaped wire would end up **1b. Draw how the coathanger would end up** **1c. Draw how the hoop would end up**

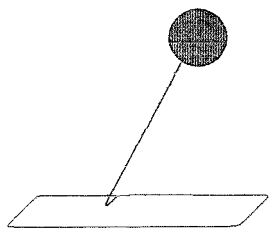
1d. Why is it in the position you have drawn?
1e. Pick up the object by the the plasticine ball and draw how you think it would end up.
1f. When we hang the objects, do they end up as you drew them? If necessary draw the hanging object again:

1g. L-shape **1h. Coathanger** **1i. Hoop**

1j. Look at the thread to which the hoop is attached. What can you say about its position?
1k. What keeps the thread in this position?
1l. When we hang the three objects together, look at the different positions of the objects. Have they got anything in common?

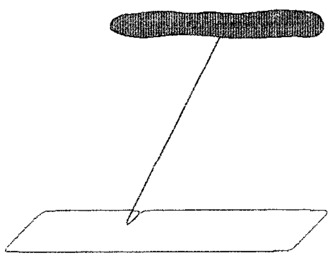
Task 2: equilibrium of objects on supporting structure

2a. Let's start by observing how the ball doesn't fall in this position. How much further can we bend the wire towards the right without the structure falling?[1]



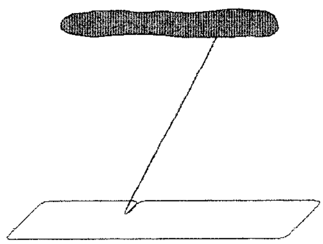
How far can it lean?

2b. Let's start with the plasticine ball about to fall (limit position). Now I'll change the plasticine ball into a sausage and put it horizontally with the wire in the centre. Will it fall? How far can we bend the wire towards the right without the plasticine falling?[1]



The wire is in the centre of the plasticine

2c. The interviewer puts the plasticine in the position (about 1/4 of way along its length) shown in 2c, where it is about to fall. The interviewer asks: Will it fall now? How far can we bend the wire to the right before the plasticine falls?[1]



The wire is ¼ of the way along the plasticine

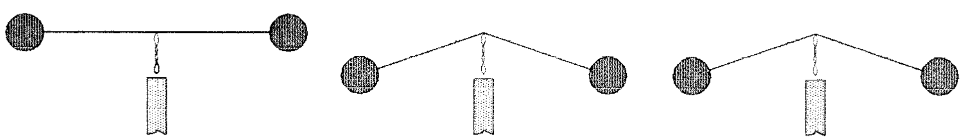
2d. The interviewer puts the plasticine sausage in a vertical position in which it will not fall and asks: How far to the right can we bend the wire before the plasticine falls?[1]

Note

1. The interviewer bends the wire with the plasticine with his/her hand on the base of the structure so that the student does not know whether the plasticine is balanced or about to fall. According to the dynamic of the interview, the student's prediction is always sought before removing the hand and an explanation after removing the hand. The initial views of the student are contrasted with his/her explanations after the experiment.

Task 3: equilibrium of biweight position

3a. Is there any position in which the balls (biweight) will not fall or will they all fall?



Position 1 **Position 2** **Position 3**

3b. All the positions are tested (1 and 2 fall, 3 does not). The student is asked why two remain balanced and the other no. *The interviewer confronts the student with the observed positions and the replies given in 3a.*

3c. After the demonstration, the interviewer concentrates the student's attention on the biweight that does not fall, asking why it remains balanced. *Free conversation.*

3d. Starting from position 1 the interviewer gradually bends the wires downwards, asking for the least degree of bending whereby the biweight does not fall. *Free conversation.*

3e. The wires are gradually straightened out so that in some cases the centre of mass is above the fulcrum. *After the demonstration, the interviewer confronts the student with the initial replies and asks for explanations. The relevant factors involved in equilibrium are not explained, but the student is asked to speak of them implicitly.*

3f. Position 3 (equilibrium). What would happen if the balls were made heavier? *Free conversation.*

3g. The balls are made heavier by adding more plasticine. The student is asked to compare his/her answer (3f) with the observed facts.

3h. Position 3 (equilibrium). Will the wires remain balanced if we remove the balls? *Free conversation.*

3i. The balls are removed and the student's initial reply is contrasted with the observed result. *Free conversation.*