ANALYSING COSTA RICAN AND SPANISH STUDENTS' COMPARISONS OF PROBABILITIES AND RATIOS

CARMEN BATANERO
University of Granada, Spain
batanero@ugr.es

LUIS A. HERNÁNDEZ-SOLÍS
State Distance University, Costa Rica
lhernandez@uned.ac.cr

MARÍA M. GEA
University of Granada, Spain
mmgea@ugr.es

ABSTRACT

We present an exploratory study of Costa Rican and Spanish students' (11–16-year-olds) competence to compare probabilities in urns and compare ratios in mixture problems. A sample of 704 students in Grades 6 through to Grade 10, 292 from Costa Rica and 412 from Spain, were given one of two forms of a questionnaire with three probability comparison and three ratio comparison problems each. The full questionnaire consisting of both forms covers six different proportional reasoning levels for each type of problem. We analysed the percentages of correct responses to the items and the strategies used by students in each grade and in each country. The results suggest students had the highest difficulty in comparing probabilities for items at the same proportional reasoning level. The results also point to the probabilistic biases in students' responses, although these biases were less frequent than in previous research.

Keywords: *Statistics education research; Comparing probabilities; Proportional reasoning; Reasoning levels; 11–16-year-old students*

1. INTRODUCTION

Probability is currently taught in primary and secondary education in countries such as Costa Rica (Ministerio de Educación Pública [MEP], 2012) and Spain (Ministerio de Educación y Formación Profesional [MEFP], 2022a, 2022b). The aim is to provide students with a strong probabilistic culture (Gal, 2005) to understand random phenomena and to make appropriate decisions in uncertain situations. This need became evident during the COVID-19 pandemic, when discussions of probability models to assess the situation and the potential risks associated with different policy decisions were prevalent in the media (Gal & Geiger, 2022; Muñiz-Rodríguez et al., 2020).

The achievement of students' learning requires a prior assessment of their abilities to cope with the tasks proposed to them in the classroom. In this paper, we focus on probability comparison, which received considerable interest in early studies on the topic in a period when most participants in that research had received no formal education in probability (Bryant & Nunes, 2012; Hernández-Solís et al., 2023; Nikiforidou, 2018; Pratt & Kazak, 2018). These investigations suggested a strong relationship between students' success with probability comparison and their proportional reasoning. Few studies, however, examined both competencies systematically and simultaneously.

This research is part of a funded research project aimed at relating Spanish students' solving of proportional problems with their solving of probabilistic problems. In this paper, we describe an exploratory assessment of 11–16-year-old-students' competence with comparing probabilities in relation to the proportional reasoning level required for comparison—a topic where more research is needed (Batanero, & Álvarez-Arroyo, 2023). Our work complements previous studies that did not systematically consider all proportional reasoning levels in probability comparison tasks for students at these ages. The sample of Spanish students is complemented with students from Costa Rica, where research in probability understanding is very scarce, and the topic was reinforced in the latest curriculum (MEP, 2012). Specifically, we intend to answer the following research questions:

1. For each grade level and each country, what is the relative difficulty of comparing probabilities in urns for problems with different proportional reasoning levels according to Noelting (1980a, 1980b)? Does this difficulty coincide with that of comparing ratios in problems with the same reasoning level?
2. Which strategies do students in the different grades from both countries use to compare ratios and probabilities? Are there incorrect strategies that are specific to the comparison of probabilities?

The answers to these questions provide new research results, which have the potential to help teachers in Costa Rica and Spain to select the adequate level of tasks with which students should work in the different grades. At the same time, teachers from other countries with similar curricula can also benefit from this information.

2. THEORETICAL FRAMEWORK

Students' performance in probability comparison tasks was first studied by Piaget and Inhelder (1951), who described different levels of reasoning in these tasks following their theory of evolutionary development. Later, Noelting (1980a; 1980b) revised the stages proposed by Piaget and Inhelder (1951) for comparison of probabilities and extended the stages to proportionality problems. Noelting's levels are considered in this section for use in this research.

2.1. RESEARCH ON STUDENTS' COMPARISONS OF TWO PROBABILITIES

Piaget and Inhelder (1951) suggested that knowledge progresses in developmental stages with an established order, although the age at which these stages are reached may vary. Subjects in the same stage exhibit similar reasoning.

To study children's reasoning when comparing probabilities, Piaget and Inhelder (1951) used a small number of white tokens, marked with a cross or unmarked, which they placed in transparent boxes. They asked the children to choose which of two boxes was preferable for obtaining a marked token. The authors changed the number of marked tokens (favourable cases) and unmarked tokens (unfavourable cases) in both boxes and conducted interviews with children from the age of 3.5 to 14 years. By comparing similar responses from groups of children of the same age, Piaget and Inhelder (1951) described the following stages in the development of children's reasoning in the comparison of probabilities:

- Stage I with two sublevels:
  - Level IA: Children do not understand the inclusion of a part in a whole, the disjunction between two types of elements (favourable and unfavourable cases), or the conservation of quantities.
  - Level IB: Children compare only one kind of token (favourable or unfavourable case) and cannot conceive of favourable cases as part of the possible cases (i.e., they do not perform a part-whole comparison). Children begin to understand that probability depends on the number of favourable or unfavourable cases.
- Stage II with two sublevels:
  - Level IIA: Children can compare probabilities only when their numbers of favourable or unfavourable cases are identical. Children use additive comparisons (e.g., subtracting the number of favourable cases from the number of unfavourable cases, or vice versa, in each box and comparing the differences). At this stage, children do not understand fraction or proportion.
  - Level IIB: Children start solving the problem when the composition of the boxes is proportional. To do so, they establish a ratio between the favourable and unfavourable cases in one box and compare it with the ratio in the other box.

- Stage III: Children can compare probabilities when they have to use proportional reasoning and can think of a general strategy if the number of favourable and unfavourable cases are small and the ratio between them is simple. Their solution becomes more general with age as the child acquires sufficient knowledge of fractions.

2.2. PROPORTIONAL REASONING AND STRATEGIES IN COMPARING RATIOS

There is extensive research on proportional reasoning due to its role in geometry, measurement, algebra, and statistics (synthesised in Ben-Chaim et al., 2012; Carpenter et al., 2012; Lamon, 2007; Resnick & Singer, 1993; Van Dooren et al., 2018). These studies have centred on various meanings of rational numbers and different tasks (Burgos & Godino, 2020). The focus of this study is the comparison of ratios, which is closely related to the comparison of probabilities. Ratio comparison served to study the development of proportional reasoning in previous studies (e.g., Karplus et al., 1983; Noelting, 1980a, 1980b). Noelting, for example, posed the problem of mixing water and lemon juice (see the items in the questionnaire forms in the Appendix). In his analysis, he expressed each mixture as an ordered pair (a_1, b_1), (a_2, b_2), where the first term in each pair (a_1 and a_2) corresponds to the number of glasses of lemon juice and the second term corresponds to the number of glasses of water (b_1 and b_2). In his work, he considered the following levels, which we employ in our reseach:

- *Level IA, Lower intuitive:* The second terms of the ratios are identical, $\{a_1 < a_2; b_1 = b_2\}$. An example is Item 1 in Form A of the questionnaire. The usual strategy is to compare the first terms (a_1 and a_2) without considering the second terms.
- *Level IB, Medium intuitive:* The first terms of the ratios are equal, $\{a_1 = a_2; b_1 > b_2\}$. Children compare the second terms and perceive them to be reciprocals of the first terms. Item 2 in Form B of the questionnaire is an example.
- *Level IIA, Lower concrete operational:* The two terms in each ratio are identical, $\{a_1 = b_1; a_2 = b_2\}$ or $\{a_1/b_1 = a_2/b_2 = 1\}$. This level is characterised by the unit equivalence class because both ratios are equal to 1. (See Item 3 in Form A.) To solve these problems, students need to use the multiplicative relations between the terms in each ratio.
- *Level IIB, Higher concrete operational or fraction equivalence class:* In this case, the terms of both ratios are proportional, $\{a_1/b_1 = a_2/b_2\}$. The difference between IIA and IIB is that the ratios are equivalent to an integer number different from unity. (See Item 4 in Form B.)
- *Level IIIA, Lower formal operational:* The terms in only one ratio are proportional, $\{m \cdot b_1 = b_2; m \cdot a_1 < a_2\}$ or $\{m \cdot a_1 = a_2; m \cdot b_1 > b_2\}$ (Item 5, Form A). To solve these problems, students need to combine multiplication and addition. They first find an integer number equivalent to one ratio and afterwards compare this value with the other ratio.
- *Level IIIB, Higher formal operational:* There is no multiplicative relationship between terms. Students compare any fractions (for example, Item 6 in Form B). The solution requires transforming the ratios to a common denominator and then comparing the first terms.

In summary, Noelting (1980a; 1980b) extended and varied slightly the levels described by Piaget and Inhelder (1951) for comparing probabilities to the comparison of ratios. In our research, we use the levels defined by Noelting (1980a;1980b), which also apply to the comparison of two probabilities when the first term of the ratio is the number of favourable cases, and the second term is the number of unfavourable cases. In Table 1 (Section 3) we summarise the types of ratios involved in each level used in our questionnaire and the typical strategy for comparing probabilities and ratios in each level.

3. RELATED RESEARCH

Piaget and Inhelder's (1951) research on probability comparison inspired a series of papers described in Bryant and Nunes (2012) and Hernández-Solís et al. (2023). Although this type of task, which is popular in research on students' understanding of probabilities, may seem easy, Supply et al. (2020) reported that in a similar PISA item, only 27% of German 15-year-old students gave the correct answer.

There is much research about comparing probabilities that involves school children. For example, Falk et al. (1980) asked 61 Israeli children (4–11-year-olds) to compare probabilities using urns with balls, and roulette wheels. They considered the following tasks: a) The number of favourable cases was smaller, higher, or identical in the higher probability group; b) The number of favourable cases was smaller, higher, or identical in the lower probability group; and c) both groups were equally likely, and the number of favourable cases was smaller, higher, or identical in each group. A systematic error was choosing the group with more favourable cases.

Table 1. Reasoning levels for comparing ratios and comparing probabilities (Noelting, 1980a, 1980b)

			Typical strategy	
Level	Description	Type of ratios	Comparing probabilities	Comparing ratios
IA	Lower intuitive	Second terms identical	Comparing favourable cases	Comparing first terms
IB	Medium intuitive	First terms identical	Comparing unfavourable cases	Comparing second terms
IIA	Lower concrete operational	Identical terms in each ratio	Recognising equiprobability	Recognising equivalence to unit
IIB	Upper concrete operational	Four terms proportional	Recognising equiprobability	Recognising equivalent ratios
IIIA	Lower formal operational	Proportional terms in each ratio	Comparing ratios between favourable and unfavourable cases	Comparing both ratios
IIIB	Upper formal operational	Any ratios	Reducing to common denominator	Reducing to common denominator

Supply et al. (2020) analysed how 177 children from Kindergarten and Grades 1–3 (5–9-year-olds) compared probabilities in two urns. The authors included items in which one urn provided only favourable cases, items with three different colours, and items with a lower number of favourable cases. Children in Grades 2 and 3 performed better than those in Grade 1 and Kindergarten. The authors found that younger children primarily decided that the urn with greatest probability of selecting a favourable case was the one with the greatest number of favourable cases. Older children realized that unfavourable cases also matter in probabilistic comparisons.

In previous research (Hernández-Solís et al., 2021), we analysed the way in which Grade 6 primary school Costa Rican children compared probabilities in urns. We did not, however, include items assessing Noelting's Level IIIB and did not use ratio comparison tasks. Results suggested that few children solved the Level IIIA items and exhibited some reasoning biases. Because the sample was small and the age range of the children was limited, we decided to investigate these results with a wider and more varied sample and relate the results to the students' level of proportional reasoning.

Research with older students is sparse, with few studies linking comparing probabilities with proportional reasoning. Green (1982) assessed 11–16-year-old English students' competence in probability with a questionnaire based on Piaget and Inhelder's (1951) experiments. The authors found the following strategies for determining the urn with greater probability for selecting a favourable case: a) choosing the urn with more possible cases; b) selecting the urn with more favourable cases; c) preferring the urn that provided the highest difference between favourable and unfavourable cases; d) choosing the urn with the highest proportion between favourable and unfavourable cases.

Cañizares et al. (1997) proposed probability comparison tasks corresponding with Levels IA to IIIB described by Noelting (1980a; 1980b) to 143 10–14-year-old students. The authors classified the children's strategies into one- and two-variable procedures. In the one-variable approaches, only favourable, unfavourable, or possible cases were compared. In two variable procedures, favourable and possible cases were combined in an additive or multiplicative way. The authors estimated the children's level of proportional reasoning according to the number of items children answered, although they did not differentiate children's reasoning level by grade. The authors obtained the following distribution: 15 students reasoned at Level 0, 29 at Level IA, 32 at Level IB, 20 at Level IIA, 24 at Level IIB, 12 at Level IIIA, and 2 at Level IIIB.

The work most closely related to the work reported in this article is that by Pérez-Echeverría et al. (1986), who used 10 ratio comparison tasks and 10 probability comparison tasks similar to the tasks used in this study. They posed those problems to 20 students aged 12 and 20 students aged 17–18. The authors classified the tasks into four levels of difficulty, according to the solving strategy required:

- Level 1: Problems where the number of favourable or unfavourable cases in both urns is identical; therefore, the solution does not require the use of fractions. This level includes Noelting's (1980a; 1980b) Level IA and IB items.
- Level 2: Problems where proportionality between the favourable and unfavourable cases is the same in both urns. These problems can be solved by establishing a ratio in one group and observing that the ratio is the same in the other. This level corresponds to Noelting's (1980a; 1980b) Level IIA and IIB items.
- Level 3: Problems with proportionality only between the favourable or unfavourable cases of both groups, or between favourable and unfavourable cases of only one group. After the ratio between favourable or unfavourable cases in one urn is perceived, it can be compared with the ratio between terms in the other urn. This level is equivalent to Level IIIA in Noelting (1980a; 1980b).
- Level 4: Problems with no proportionality relationship between the four members (favourable and unfavourable cases of each grouping). These problems require working with fractions and transforming them to a common denominator. Level 4 is equivalent to Level IIIB in Noelting (1980a; 1980b).

The percentage of multiplicative strategies used by children was only 12% in proportionality tasks and 10% in probability tasks (Pérez-Echeverría et al., 1986). Furthermore, the participants who used a proportional strategy to solve Level 2 or Level 3 problems reverted to additive strategies or strategies for comparing absolute quantities to solve the Level 4 tasks. The use of correspondence strategies was higher in older students (50%). Overall, the theoretical difficulty levels corresponded to the observed difficulty levels as the percentage of incorrect solutions increased with the task level (65% of subjects in the ratio problem and 75% in the probability problem at the highest-level provided incorrect responses).

Following that work, Berrocal (1990) presented three tasks (a probability task, a proportionality task, and a covariation task) to 103 students from Grade 7 (53 students) and at university level (50 students). Berrocal only reported the correlation between the results of the three tasks. In a second experiment using the same tasks with 305 children aged 11 to 15 years, she examined the task difficulty, informing that the probability task was the most difficult (Berrocal, 1990).

These studies with Spanish students used only small sample sizes, students not representative of the entire age range in which the development of the concept unfolds, or a few isolated tasks (Berrocal, 1990). None of the researchers reported in detail the correct and incorrect strategies at different ages. To complement these studies, we carried out an exploratory analysis of 11–16-year-old students' performance with probability comparisons and ratio comparisons and their associated strategies. We compared the performances and strategies of Spanish students with the performances and strategies of age-matched Costa Rican students.

4. METHODS

4.1. SAMPLE

The sample consisted of 704 students, 292 Costa Ricans and 412 Spaniards, from the last grade of primary education (Grade 6 in both countries¸ 11–12-year-olds), the four grades of Spanish Compulsory Secondary Education (CSE) (12–16-year-olds), and equivalent grades (Basic Education Grades 7–9 and Diversified Cycle Grade 10) in Costa Rica. In Table 2 we present The sample composition of students who completed each of the two forms of the questionnaire was 45.3% girls, 50.8% boys; the remaining students did not report their gender. Because we took full school groups, the number of students vary for both grades and countries.

Table 2. Sample composition according to grade, country, and questionnaire form

	Form A		Form B		Total	
Grade	Spain	Costa Rica	Spain	Costa Rica	Spain	Costa Rica
6	39	35	39	33	78	68
7	43	26	37	26	80	52
8	32	31	31	33	63	64
9	50	26	46	26	96	52
10	55	27	40	29	95	56
Total	219	145	193	147	412	292

In Spain, four schools participated in the study: two primary schools (from Almería and Granada) and two secondary schools (Almería and Seville). The sample of Costa Rican students was taken from a single primary and secondary school in the province of Cartago. All the schools included students from different social and economic levels and were similar to most other schools in their respective countries. In each school, two or three complete groups of students participated per grade (depending on the school). In total, we included 13 groups in Costa Rica and 24 groups in Spain.

This sample constituted a convenience sample for which the subjects in the sample were easily accessible; geographical proximity; availability at the time of the study; and willingness to participate. Convenience samples are often used in educational research and are reasonable to use when the population composition is homogeneous and the aim is not to generalize to the entire population (Etikan et al., 2016).

When the questionnaire forms were administered, the students had studied probability and proportionality according to the guidelines in effect in their respective countries. Both countries introduce probability at the beginning of primary education using games of chance and problems in familiar contexts (MECD, 2014; 2015; MEP, 2012). In the last year of primary education (Grade 6), students should use the intuitive ideas previously acquired to compute probability by the Laplace rule in Costa Rica. In Spain, children study this rule in Grade 7, whereas there is no probability in this grade in Costa Rica. In Grade 8, all the probability ideas introduced in previous grades are revisited, with increasing difficulty of problems. In Grade 9 in Costa Rica and Grade 10 in Spain, the frequentist definition of probability and the law of large numbers are introduced. In Grade 10, the properties of union and complement are studied in Costa Rica, and conditional probability is studied in Spain.

Rational numbers are studied in Spain (MECD, 2014; 2015) and Costa Rica (MEP, 2012). In primary education, the concept of fractions and the reading, writing, representation, ordering, and basic operations are introduced to later differentiate between proper and improper fractions. Equivalent fractions and percentages are also introduced in situations of direct proportionality to solve everyday problems. In Grades 7 and 8 in Spain and Grade 8 in Costa Rica, fractional numbers and calculation strategies are used to simplify operations, and inverse proportionality is introduced. The concept of a rational number and its properties are studied in Grade 8 in Costa Rica and Grade 9 in Spain. There is no new content associated with proportional reasoning introduced in the following years, although proportional reasoning is used in the resolution of problems.

4.2. QUESTIONNAIRE

We built a questionnaire containing six ratio comparison problems similar to those used by Noelting (1980a; 1980b), Karplus et al. (1983), and Tourniaire and Pulos (1985) and six probability comparison in urns problems with increasing proportional reasoning levels that resemble those used by Cañizares et al. (1997), Green (1982), and Pérez-Echeverría et al. (1986). The questionnaire, divided into two forms, A and B, is included in the Appendix. Six reasoning levels by Noelting (1980a; 1980b) appear in the questionnaire (Table 3). We decided to give each student only half of the questionnaire so as not to over-fatigue students and to avoid response bias as the questionnaire progressed. Each form contains the same reasoning levels for ratio and probability items. We gave each form in a randomized way to half of the students in each group of students, school grade, and country to ensure that an approximately equal number of students, selected randomly, answered each item.

Table 3. Reasoning levels (Noelting, 1980a, 1980b) required to solve the items

Item	Item type	Composition (a_1, b_1) vs (a_2, b_2)	Level (Noelting)	Proportional Reasoning Level	Questionnaire Form
1	Comparing ratios	(2, 3) vs (1, 3)	IA	Lower intuitive	A
2	Comparing ratios	(5, 1) vs (5, 4)	IB	Medium intuitive	B
3	Comparing ratios	(2, 2) vs (4, 4)	IIA	Lower concrete operational	A
4	Comparing ratios	(3, 1) vs (6, 2)	IIB	Upper concrete operational	B
5	Comparing ratios	(3, 1) vs (4, 2)	IIIA	Lower formal operational	A
6	Comparing ratios	(3, 2) vs (4, 3)	IIIB	Upper formal operational	B
7	Comparing probabilities	(3, 2) vs (5, 2)	IA	Lower intuitive	A
8	Comparing probabilities	(4, 1) vs (4, 3)	IB	Medium intuitive	B
9	Comparing probabilities	(2, 2) vs (3, 3)	IIA	Lower concrete operational	A
10	Comparing probabilities	(2, 6) vs (1, 3)	IIB	Upper concrete operational	B
11	Comparing probabilities	(3, 6) vs (1, 3)	IIIA	Lower formal operational	A
12	Comparing probabilities	(3, 4) vs (4, 5)	IIIB	Upper formal operational	B

We firstly determined the questionnaire's proposed use and interpretation: to compare the students' performance and strategies in two types of items at the different proportional reasoning levels of Noelting (1980a, 1908b): a) comparing two ratios in the context of mixtures and b) comparing two probabilities in the urns context. For each of these two types of items, we compiled an item bank with tasks taken from the research described in Section 3.

For each questionnaire item, two ratios or probabilities (a_1, b_1) and (a_2, b_2) are compared. The first term of each pair is the antecedent or dividend of the ratio (glasses of lemon juice/glasses of water), and the second term is the consequent or divisor in the comparison of ratios. In the comparison of probabilities, the terms represent the number of favourable and unfavourable cases. For example, in Item 1 from Form A, the ratios are 2/3 and 1/3 and can be solved by only comparing the antecedents because the consequents are identical. Likewise, Item 7 from Form A can be solved by comparing the favourable cases. Both items are of the same proportional reasoning level (Lower intuitive).

The construction of the questionnaire was rigorous, following the *Standards for Educational and Psychological Testing* published by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) to assure validity and reliability. The *Standards* describe validity as a unitary concept "implying the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use" (p. 14). We compiled validity evidence in different ways.

Validity evidence based on the questionnaire content (AERA, APA, & NCME, 2014, p. 14) was firstly obtained from the analysis of the relationship between that content and the construct we intended to measure (Item type, Table 3). Following these standards, test content refers to the themes, wording, and format of the items, tasks, or questions on a test (p. 14). Secondly, we assured further *validity evidence based on content* from experts' judgments (AERA, APA, & NCME, 2014, p. 14). A questionnaire containing three different possible versions of each item was sent to ten different researchers who are experts in statistics education. They scored the items using a five-point Likert scale for the adequacy of each version for each item. The final items included in the questionnaire were those with the highest median and mean scores (4 points or more) and the least variability in the experts' scoring.

According to the *Standards* (AERA, APA, & NCME, 2014, p. 64), instrument developers should include relevant subgroups in preliminary item and instrument evaluation studies when constructing the instrument. With this purpose, and *to obtain new evidence of validity*, the final questionnaire was trialled in individual interviews with ten children aged 10–12 to check their understanding and to confirm the questionnaire's validity—these children were not included in the final sample for the study.

The children who were interviewed perceived the difference between both types of items as seen in the following responses:

M (age 11): When you prepare the lemonade many times you always get the same flavour, it is certain, in the games of picking up a counter you can obtain different results, …, even if black is more likely, you can get a white counter by chance. It is a matter of luck.
S (age 12): They are similar; but in the first exercise you should use percentages and in this (the urn item) you need the logic.

They also perceived the chance element in the urn problems when asked if they will always obtain the same result when repeating an experiment, as expressed by M and in the next examples:

L (age 12): There is nothing to assure you always will get black in box B; but there are more possibilities because of the number of black balls.
S (age 12): Almost always you will pick a black ball, because the probabilities are higher 80% and only 20% of white balls.

Additionally, we obtained reliability evidence. Thus, the Cronbach's alpha reliability coefficient of the questionnaire in the participant sample was $\alpha = 0.738$ for Form A and $\alpha = 0.734$ for Form B. These values are reasonable, given that the questionnaire measures two types of tasks (comparing ratios and comparing probabilities).

4.3. STRATEGY CATEGORIES

After we collected the questionnaires, we performed a content analysis of the responses to each item. This method permits the establishment of or refinement of categories that emerge objectively as a result of systematic study (Krippendorff, 2013). We first divided the strategies into categories of (mathematically) correct and incorrect. The strategies are correct if they provide a correct solution to the problem and incorrect if the solutions are not adequate for solving the problem. Secondly, we started from the strategies defined in previous research such as Cañizares et al. (1997), Green (1982), and Noelting (1980a, 1980b). Using an inductive and cyclical process, including successive revision and discussion of cases, we refined the classifications. Responses were coded by one author. Another author coded the responses in a subsample of 20 students (24 variables each; in total 480 codes) to compute the inter-coder reliability. We obtained values of Cohen's kappa: $\kappa = 0.97$ for the coding of responses and $\kappa = 0.92$ for the coding of strategies. These values are very high because $\kappa > 0.8$ is interpreted as almost correct agreement (Cohen, 1960). Below we describe the different strategies with examples of students' responses in each category. Students are denoted by S*x*, where *x* is the student's order in the data file. The strategies were deduced from students' justifications and the options selected by them. For the probability items, these options were: a) You are more likely to draw a black counter from box A; b) You are more likely to draw a black counter from box B; or c) There is an equal chance of drawing a black counter in both boxes.

Comparing totals. The strategy involves comparing the sum of terms in both ratios (comparing the possible cases in both urns). This strategy is incorrect for all the items and was described by Green (1982).

S205: Because she has fewer glasses (Item 4, option A).
S1: Because in Box B there are fewer counters. Therefore, it will be easier to pick a black counter than in Box A where there are more counters, Box B is more likely (Item 11, option A).

Comparing the first terms of the ratios (comparing favourable events). This procedure provides correct results only when the second terms (unfavourable cases) are the same ($b_1 = b_2$). Students already at Noelting's lower intuitive level (IA) can use it. The student must differentiate both terms of the ratio. In this study, the strategy provides correct answers to item 1 (S346) and item 7 (S8). Some students, such as S301 and S42, applied the method incorrectly to other problems.

S346: Elena, because she got more juice (Item 1, option A).
S8: Because there are more black balls in A than in B (Item 7, option A).
S301: Because there are more lemon glasses (Item 5, option B).
S42: Because in box B there are more black balls than in box A; therefore, B is more likely (Item 9, option B).

Comparing the second terms of ratios (comparing unfavourable events). This procedure provides correct answers if the first terms are identical ($a_1 = a_2$). The student must understand that with the same numerator (same number of favourable cases), the ratio is larger (probability is higher) if the denominator (number of unfavourable cases) is smaller. This strategy corresponds to Noelting's intermediate medium level (IB). The strategy gives correct answers to item 2 and item 8, although some students used it incorrectly (see S25).

S172: Because Elena has one glass of water and Juan has four (Item 2, option A).
S54: Because there are more white counters in box B (Item 8 option A).
S198: Because there is less water (Item 5, option A).
S25: Because there is the same number of white balls (Item 7, option C).

Comparing the differences between the terms of each ratio (comparing the differences between favourable and unfavourable cases). This strategy involves considering the four terms in the problem but making additive comparisons by analysing the difference between these terms. According to Noelting (1980a; 1980b), this strategy indicates that the student constructs the ratio. The student perceives the difference between terms in each ratio and selects the one with the higher difference. The strategy is appropriate for solving Items 1, 2, 7, and 8 of the questionnaires, but was used incorrectly by some students (e.g., S197 and S655).

S39: Because Elena has one more glass of water than juice and Juan has two more glasses of water than juice. (Item 1, option A).
S1: A black counter is more likely in B because there are more black counters than white counters (Item 7, option B).
S197: Because Juan got one more glass of lemon and one more glass of water (Item 6, option C).
S655: Because there is only a difference of 1 between the two boxes (Item 12, option C).

Ratio of equivalence to unity. The student compares one ratio (a_1/b_1) with the other (a_2/b_2), that is, applies a multiplicative operation to the terms in both ratios, discovering an equivalence to unity (discovering equiprobability in probability comparison tasks). Although the previously described strategies involve comparing the terms of each ratio or their differences, in this case the terms of both ratios should be considered. The strategy provides correct answers when the terms of each ratio are identical. This strategy corresponds to Noelting's lower concrete operational level (IIA). In the study, the strategy gives appropriate answers for Item 3 and Item 9 and was always used correctly by students.

S421: Because both lemonades have the same amount of juice and water in different quantities (Item 3, option C).
S446: In both urns there are the same amount of black and white counters (Item 9, option C).

Equivalence relation between ratios. The student compares one ratio with the other using a multiplicative operation, finding equivalence of both ratios. This strategy leads to correct answers only when the ratios belong to the same equivalence class of fractions. The procedure corresponds to Noelting's higher concrete operational level (IIB) and gives correct answers in item 4 and item 10. The strategy was misapplied in some cases (e.g., S212 and S134).

S317: Because Elena's amount of lemonade is smaller and so is the quantity of water, those of Juan are multiplied by 2 (Item 4, option C).
S14: Because in urn A there is a third as many white counters as black counters and the same for urn B (Item 10, option C).
S212 Because both mixtures have the same proportion (Item 6, option C).
S134: In both boxes there are half black balls (Item 11, option C).

Correspondence between the terms of each ratio. The student applies a proportionality criterion between the terms of each ratio to determine which of these proportions is smaller. This strategy gives correct answers if the terms of the two ratios are multiples. It corresponds to Noelting's lower formal operational Level IIIA and can be used to solve Item 5 and Item 11. The strategy usually was applied correctly, with a few failures in students who established incorrect correspondences (e.g., S337, S291).

S506: It is concentrated in less water. For each glass there are 3 glasses of lemon, so Juan should use 6 glasses of lemon to be as acid as Elena (Item 5, option A).
S17: Because urn A has half as many black counters as there are white counters and urn B has one third as many black counters as there are white counters (Item 11, option A).
S337: That from Elena, because there is 1/3 of lemonade, while Juan got ¼ of lemonade (Item 6, option A).
S291: In urn A the balls are ¾, while in urn B they are 4/5 (Item 3, option C).

Proportionality. Ratios are reduced to common denominator fractions and compared. With this strategy, students compare ratios and give correct answers for any ratio comparison task. It corresponds to Noelting's higher formal operational Level IIIB. The strategy can be used to solve all questionnaire items.

S80: Because the percentage of lemon juice in Elena's lemonade (60%) is higher than that of Juan's (57%) (Item 6, option A).
S303: The probability of getting a black counter is 50% in both urns, since the number of counters is the same for both colours (Item 9, option C).

Biased strategies. Some students provided strategies in the comparison of probabilities that suggested they hold probabilistic biases. Such incorrect reasoning included the equiprobability bias (Lecoutre & Durand, 1988), the outcome approach (Konold, 1989), and judging probability by the physical properties of the random generator.

Equiprobability. In the following example, S614 suggests equiprobability in both urns in Item 7, although the student is conscious that in one urn there are more favourable cases than in the other.

S614: In urn B, obviously there are more black counters; but it is possible to get a black counter in both urns, so the probability is the same (Item 7, Option C).

Outcome approach. Other students did not interpret the problem probabilistically. Instead, they understood that they should predict the outcome of the random experiment. This interpretation is incorrect because random events are unpredictable, although we can estimate their probability. This reasoning was termed the *outcome approach* by Konold (1989).

S10: There may be different results because you pick it with closed eyes. Even if the urn with larger number is more likely, this does not influence the results (Item 9, Option C).

Physical considerations. A few students referred to physical considerations. For example, S103 assumed that a small movement in the urns would affect the results due to the initial position of counters.

S103: Because when we shake the urn, the white ball will rise to the top in urn A, but there is still a chance of getting a black ball. When we shake urn B, the black balls will rise to the top, but there is still a probability that the white ball will come out (Item 8, option C).

Unclear. Finally, some strategies were unclear and other students did not explain their procedure.

5. RESULTS

In this section, we present the results for each group of research questions posed in the introduction.

5.1. CORRECT RESPONSES BY ITEM

In this section we answer the first set of research questions: For each grade level and each country, what is the relative difficulty of comparing probabilities in urns for problems with different proportional reasoning levels according to Noelting (1980a, 1980b)? Does this difficulty coincide with that of comparing ratios in problems with the same reasoning level? To provide an answer to these questions, in Table 4 we produce the percentage of correct responses for comparison of probability items. In Table 5 we report the percentage of correct responses for comparison of ratio items.

Table 4. Percentage of correct answers in comparison of probability items by country and grade

		Grade									
		6		7		8		9		10	
Item	Noelting level	Spain	CR	Spain	CR	Spain	CR	Spain	CR	Spain	CR
7	IA	79.5	85.7	86.0	88.5	84.4	93.5	80.0	88.5	80.0	81.5
8	IB	92.3	93.9	86.5	80.8	90.3	87.9	93.5	84.6	87.5	96.6
9	IIA	53.8	51.4	72.1	69.2	96.9	61.3	84.0	80.8	81.8	74.1
10	IIB	17.9	18.2	35.1	23.1	45.2	24.2	37.0	26.9	65.0	41.4
11	IIIA	35.9	45.7	30.2	42.3	59.4	48.4	44.0	61.5	56.4	51.9
12	IIIB	33.3	45.5	35.1	30.8	9.7	33.3	26.1	42.3	42.5	31.0

Table 5. Percentage of correct answers in comparison of ratio items by country and grade

		Grade									
		6		7		8		9		10	
Item	Noelting level	Spain	CR	Spain	CR	Spain	CR	Spain	CR	Spain	CR
1	IA	89.7	91.4	88.4	92.3	93.8	90.3	86.0	92.3	94.5	96.3
2	IB	87.2	90.9	91.9	96.2	96.8	97.0	95.7	100.0	92.5	96.6
3	IIA	64.1	65.7	79.1	69.2	84.4	80.6	82.0	76.9	94.5	70.4
4	IIB	28.2	30.3	40.5	42.3	71.0	48.5	71.7	53.8	77.5	58.6
5	IIIA	41.0	40.0	46.5	57.7	59.4	48.4	36.0	50.0	58.2	48.1
6	IIIB	17.9	15.2	27.0	23.1	35.5	18.2	32.6	26.9	37.5	27.6

Tasks classified as Level IA and Level IB (lower and medium intuitive level in Noelting (1980a, 1980b) were very easy for the students who solved them correctly, which was a high percentage (over 80% of students) in grades 6–10 for both types of items. These items are solvable by comparing only the numerator or denominator of ratios (cases favourable or unfavourable in the comparison of probabilities).

The difficulty increased for level IIA problems (equivalence to unit class), with a slightly higher percentage of sixth and seventh grade students correctly responding to the comparison of probability items than the comparison of ratio items. In these items, the numerators (favourable cases) and denominators (unfavourable cases) are identical in each mixture (urn). The difficulty in each level of the comparison of ratios generally diminished with succeeding grade levels, particularly in Spain with 94.5% correct solutions to level IIA problems in Grade 10. The pattern of success for probability items was less clear. The percentages improved with increasing school years but presented oscillations. However, the percentage of students selecting the correct responses was higher in Spain (compared with Costa Rica), mainly for the probability items.

Success in Level IIB problems (equivalence class of any ratio) progressively increased by grade with more strength in comparison of ratios in Spain (compared with Costa Rica). The pattern was not clear with Level IIIA items, and Level IIIB items were too hard for students in all grades and in both countries. However, the percentage of Spanish students selecting the correct responses was slightly higher for the comparison of ratios items (compared with Costa Rica). Moreover, for the comparison of probabilities, Item 10 (IIB) was harder than Item 11 (IIIA) for Costa Rican students from all grades, and from some grades in Spain.

In general, the difficulty was slightly higher for items comparing probabilities (Items 7 to 12) than for items comparing ratios (Items 1 to 6) in both countries, except for the Level IIIB item. Yet, some students solved the Level IIB item using incorrect strategies, as described in the next section.

5.2. CORRECT STRATEGIES

To analyse which correct strategies were used by the students in the different grades and countries to compare ratios and probabilities, in Tables 6 and 7 we present the percentage of students in each grade and each country who used correct strategies to solve the different items.

Table 6. Percentage of correct strategies for comparing probabilities by item, grade, and country

		Grade									
		6		7		8		9		10	
Item (Level)	Strategy	S	CR	S	CR	S	CR	S	CR	S	CR
7(IA)	Comparing favourable cases	53.8	48.6	60.5	57.7	28.1	54.8	44.0	57.7	30.9	44.4
	Comparing differences	25.6	34.3	34.9	30.8	34.4	38.7	36.0	26.9	32.7	29.6
	Correspondence					3.1					
	Proportionality		2.9		3.8	18.8		2.0	3.8	16.4	3.7
8(IB)	Comparing unfavourable cases	38.5	42.4	32.4	30.8	26.7	27.3	28.3	34.6	27.5	31.0
	Comparing differences	43.6	39.4	37.8	34.6	53.3	27.3	41.3	26.9	22.5	31.0
	Correspondence						3.0			2.5	
	Proportionality	2.6		8.1			9.1	8.7	7.7	17.5	3.4
9(IIA)	Equivalence to unity	38.5	48.6	60.5	57.7	68.8	54.8	58.0	57.7	56.4	59.3
	Proportionality	2.6	5.7	11.6	11.5	21.9	9.7	12.0	7.7	27.3	11.1
10(IIB)	Equivalence to ratio	10.3	6.1	18.9	11.5	26.7	12.1	23.9	3.8	35.0	17.2
	Correspondence	2.6					6.1		3.8		3.4
	Proportionality	2.6	3.0	8.1	7.7	6.7	12.1	8.7	15.4	27.5	6.9
11(IIIA)	Correspondence	2.6	14.3	9.3		37.5	12.9	12.0	30.8	9.1	7.7
	Proportionality		5.7	4.7		21.9	12.9	4.0	3.8	32.7	7.7
12(IIIB)	Proportionality			8.1		6.5	9.1	10.9	15.4	17.5	10.3

Table 7. Percentage of correct strategies for comparing ratios by item, grade, and country

		Grade									
		6		7		8		9		10	
Item (Level)	Strategy	S	CR	S	CR	S	CR	S	CR	S	CR
1(IA)	Comparing first terms	71.8	54.3	72.1	69.4	50.0	58.0	40.0	54.0	50.9	51.9
	Comparing differences	10.3	22.9	9.3	7.7	18.8	32.3	30.0	11.5	20.0	29.6
	Correspondence				3.8	3.0					3.7
	Proportionality							4.0	11.5	18.2	3.7
2(IB)	Comparing second terms	51.3	45.5	59.5	76.9	51.7	60.6	47.8	53.8	40.0	34.5
	Comparing differences	17.9	42.4	27.0	15.4	41.9	15.2	28.3	26.9	27.5	37.9
	Correspondence						3.0				
	Proportionality							2.3	7.8	10	7
3(IIA)	Equivalence to unit	51.1	68.5	69.8	65.6	62.5	58.1	62.0	57.7	45.5	59.3
	Correspondence			7.0			3.2				
	Proportionality	2.6		2.3	3.8	12.5	3.2	10.0	11.5	40.0	11.1
4(IIB)	Equivalence to ratio	17.9	24.3	35.3	31,0	64.5	24.2	54.5	23.2	30.0	41.5
	Correspondence	5.1		2.7	3.8		6.1	2.2	3.8	2.5	
	Proportionality		3.0	2.7	3.8	3.2	9.1	8.7	15.4	37.5	10.3
5(IIIA)	Correspondence	7.7	11.4	16.3	11.6	43.7	19.3	12.0	26.9	20.1	25.9
	Proportionality	2.6			3.8	6.3		10.0	3.9	30.9	11.1
6(IIIB)	Proportionality			5.5		9.7	3.0	8.7	7.7	30.0	10.4

The high frequency of comparisons using only one term in each ratio (comparing only favourable or unfavourable cases in the probability) for Level IA and IB items where these strategies are correct, is notable. We remark that the percentage of students using these strategies was, in general, smaller in the comparison of probabilities.

The additive comparisons (differences between the first and second term of the ratio or the number of favourable and unfavourable cases in probability) are valid for items at lower levels, for example, in Items 1, 2, 7, and 8. Additive comparisons were common for these items, with higher percentages for probability items for both countries. With the comparison of differences between favourable and unfavourable cases, the students compensated for the lower use of only one variable strategies for the probability items. In items of concrete operational Levels IIA and IIB, the students changed to using equivalence to unit or equivalence to ratio procedures.

The more advanced strategies such as correspondence or proportionality appeared in responses from a smaller percentage of students in Grades 6–10, which suggests that the students in the upper grades were not familiar with these methods.

To sum up, students tended to use correct strategies for Levels IA and IB, with decreasing use of these correct strategies in higher-level items. The percentage of correct strategies diminished for the probability items; however, the students used more elaborated strategies for these items such as for the comparison of differences in Levels IA and IB. We interpret these results by the fact that the probability items were harder for the students (see Tables 4 and 5); therefore, they needed to resort to more complex strategies to obtain the correct solution.

5.3. INCORRECT STRATEGIES

To consider the incorrect strategies that students at each grade level from each country used to solve tasks, we present the percentages of students using incorrect strategies for probability items and ratio items in Tables 8 and 9, respectively. Strategies classified as unclear were from students who presented unclear strategies or no explanation for their strategies.

Table 8. Percentages of incorrect strategies for comparing probabilities by grade and country

		Grade									
		6		7		8		9		10	
Item(Level)	Strategy	S	CR	S	CR	S	CR	S	CR	S	CR
7(IA)	Comparing possible cases	2.7	5.7	2.3	7.7	3.1	3.2	2.0	7.8	5.5	7.4
	Comparing unfavourable cases	12.8	5.7			12.5	3.3	4.0		5.5	
	Equivalence to ratio									1.8	
	Biased strategies	5.1	2.9					8.0	3.8	5.4	11.2
	Unclear			2.3				4.0		1.8	3.7
8(IB)	Comparing possible cases	2.4	3.0		3.8	3.3	3.0	4.3	11.5		6.9
	Comparing favourable cases	10.3	6.1	18.9	23.1	6.6	21.2	15.2	7.7	22.5	13.9
	Biased strategies	2.6				3.3	6.1		3.9	7.5	6.9
	Unclear		9.1	2.8	7.7	6.8	3.0	2.2	7.7		6.9
9(IIA)	Comparing possible cases	7.7	14.2	2.4			9.7	4.0		7.3	7.4
	Comparing favourable cases	23.0	20.0	9.3	26.9		12.9	8.0	19.2	1.8	11.1
	Comparing unfavourable cases	10.3		2.3		3.1				1.8	
	Comparing differences	7.7	2.9	9.3	3.9	3.1	3.2	6.0	3.8	1.8	
	Biased strategies	5.1		2.3		3.1	9.7	6.0	11.6	3.6	7.4
	Unclear	5.1	8.6	2.3				6.0			3.7
10(IIB)	Comparing possible cases	17.9	27.2	10.9	15.5	33.5	15.1	21.7	42.4	2.5	27.6
	Comparing favourable cases	10.3	15.2	16.2	11.5	3.3	15.1	10.9	7.8	15.0	6.9
	Comparing unfavourable cases	33.2	18.2	32.4	26.9	13.3	27.3	15.2	11.6	10.0	13.8
	Comparing differences	12.8	24.2	8.1	23.1	10.0	6.1	6.5	3.8	2.5	20.8
	Equivalence to unit					3.3			3.8		
	Biased strategies	7.7		2.7	3.8	3.3		10.9	3.8	7.5	3.4
	Unclear	2.6	6.1	2.7			6.1	2.2	3.8		
11(IIIA)	Comparing possible cases	12.8	20.0	18.5	23.1	6.2	32.3	8.0	7.7	10.9	22.3
	Comparing favourable cases	25.6	20.0	11.6	26.9	9.4	16.1	24.0	15.4	14.5	33.3
	Comparing unfavourable cases	33.3	5.7	30.2	30.8	6.3	12.9	16.0	15.5	12.8	14.2
	Comparing differences	23.1	25.7	14.0	19.2	6.3	3.2	20.0	11.5	9.1	3.7
	Equivalence to unit						3.2				
	Equivalence to ratio			7.0				2.0			
	Biased strategies	2.6	2.9			12.4	6.5	10.0	11.5	10.9	7.4
	Unclear		5.7	4.7				4.0	3.8		3.7
12(IIIB)	Comparing possible cases	17.9	9.1	5.4	3.5	22.6	6.1	8.6	11.5	7.5	3.4
	Comparing favourable cases*	15.4	30.3	24.3	13.9	12.9	27.2	13.0	11.5	17.5	13.8
	Comparing unfavourable cases	7.7	9.1	10.8	10.3	3.2	9.1	6.5	7.8	2.5	10.4
	Comparing differences	30.8	45.4	43.2	13.8	41.9	33.2	17.4	38.5	12.5	48.3
	Equivalence to unit								3.8		
	Equivalence to ratio					3.2					
	Correspondence	5.1						2.2		5.0	
	Biased strategies	12.8		2.8	37.9	6.5	6.2	21.8	11.5	15.0	10.4
	Unclear	10.3	6.1	5.4	10.3	3.2	9.1	19.6		22.5	3.4

There was scarce use of each incorrect strategy for all the items. Adding the percentages for all the incorrect strategies yields a noticeable sum. In particular, the unclear strategies were more common for the comparison of ratios and specifically for the Spanish students. The most common incorrect strategies used by students were the comparison of only one term in the ratios (only favourable or only unfavourable cases in probability) as well as comparing differences, which only correctly solve Level IA and Level IB items using the classification of Noelting (1980a, 1980b). Comparing differences or first terms (favourable cases) or determining equivalence to ratio were more commonly used strategies for comparing ratios than comparing probabilities, where students tended to compare unfavourable cases more. The differences between countries were higher in the comparison of differences.

Table 9. Percentage of incorrect strategies for comparing ratios by grade and country

		Grade									
		6		7		8		9		10	
Item (Level)	Strategy	S	CR	S	CR	S	CR	S	CR	S	CR
1(IA)	Comparing totals	7.7	11.4	2.3	3.8	9.4		4	3.8	1.8	
	Comparing second terms			2.3	7.7						
	Equivalence to unit							2			
	Unclear	10.2	11.4	14	7.6	18.8	9.7	20	19.2	9	11.1
2(IB)	Comparing totals	2.6									6.9
	Comparing first terms	15.4	3	8.1	3.8	3.2	12.1	6.5	3.8	2.5	10.3
	Equivalence to unit		3								
	Unclear	12.8	6.1	5.4	3.9	3.2	9.1	15.2	7.7	20	3.4
3(IIA)	Comparing totals	7.7	2.9					2	7.7		
	Comparing first terms	15.4	11.4	9.3	7.7	6.3	16.1	8	7.7	1.8	7.4
	Comparing second terms	2.6		2.3	3.8		3.2	2			
	Comparing differences	10.3	5.7	2.3	3.8		6.5			7.3	3.7
	Unclear	10.3	11.5	7	15.3	18.7	9.7	16	15.4	5.4	18.5
4(IIB)	Comparing totals	5.1	3	5.4							6.9
	Comparing first terms	38.5	18.2	24.3	19.2	6.5	15.2	15.2	23.1	12.5	13.8
	Comparing second terms		3.0	5.4	3.8		9.1	4.3	3.8	2.5	6.9
	Comparing differences	23.1	48.5	16.2	23.1	1	24.2	4.3	19.2	5	17.2
	Unclear	10.3	0	8.1	15.3	6.4	12.2	10.8	11.5	10	3.4
5(IIIA)	Comparing totals		5.7			3.1				1.8	
	Comparing first terms	17.9	14.3	23.3	23.1	6.3	12.9	10	23.1	9.1	3.7
	Comparing second terms	20.5	8.6	18.6	15.4	3.3	9.7	2	7.7	1.8	7.4
	Comparing differences	20.5	34.3	11.6	30.8	9.4	38.7	32	19.2	21.8	29.6
	Equivalence to unit	2.6	5.7	4.6	3.8	9.4					7.4
	Unclear	28.2	20	25.6	11.5	18.7	19.4	34	19.2	14.5	14.8
6(IIIB)	Comparing totals	2.6		5.4	15.4		3				
	Comparing first terms	12.8	3	13.5	7.7		3	8.7	7.7	5	10.3
	Comparing second terms	7.7	6.1	13.5	7.7	16.1	6.1	13	15.4	7.5	6.9
	Comparing differences	46.2	72.7	35.1	42.3	45.2	51.5	34.8	46.2	27.5	44.8
	Equivalence to unit	12.9	9.1	2.7	7.7	9.7	9.1	2.2	3.8	2.5	6.9
	Correspondence		6.1		3.9	6.6	3	2.2		2.5	6.9
	Unclear	18	3	24.3	15.3	12.9	21.3	30.4	19.2	25	13.8

A few students revealed biases in their responses to the probability items such as the equiprobability (Lecoutre & Durand. 1988), outcome approach (Konold. 1989), or physical considerations. These cases were scarce; we assume the teaching of probability experienced by these students helped them to overcome the said biases.

6. DISCUSSION AND CONCLUSIONS

In this paper, we analysed the responses of a sample of Costa Rican and Spanish students (11–16-year-olds) when comparing ratios and when comparing probabilities using six items with different levels of difficulties identified by Noelting (1980a, 1980b). These results complement previous research because no related data were available in Costa Rica, and research carried out in Spain did not include the same ages and used only moderate-size samples. Moreover, we found no study analysing systematically and jointly the comparison of probabilities and ratios in the same age range.

We reported the percentages of correct responses for questionnaire items and the correct and incorrect strategies used by students in Grades 6–10 for both Costa Rica and Spain. Given that the sample of students was not random and few different schools took part, we recognise the limitations of our research. Consequently, our conclusions presented below should be taken with caution.

The percentages of correct responses revealed that the difficulty of both types of problems increased quickly for Level IIA and succeeding levels. Generally, a higher percentage of students in upper grades solved problems at all of Noelting's (1980a; 1980b) reasoning levels; however, probability problems were generally harder than the equivalent ratio problems for students from both countries. These results coincide with those obtained by Berrocal (1990), although she only used a single task without considering all of the possible reasoning levels of Noelting (1980a, 1980b). Moreover, in our sample, the students had studied probability previously, which was not true for the students in Berrocal's study. Anyway, the rate of success was high for the probability comparison tasks at Levels IA, IB, and IIA for all grades; consequently, students in all of the examined grades should be able to solve comparison of probabilities problems.

We remark that the difficulty levels in Pérez-Echeverría et al.'s (1986) research applies to comparison of ratio problems but not to comparison of probability problems at difficulty Level 3 in their classification. Item 10 (Level 2 in Pérez-Echeverría et al.'s classification) was harder than Item 11 and Item 12 (Level 3 and Level 4, respectively) for students in some grades. We attribute these differences to the smaller sample size used by Perez- Echeverría et al. (1986). In our sample, many students compared totals to solve Item 10, which is an incorrect strategy.

When compared with our previous exploratory study with Grade 6 children (Hernández-Solís et al., 2021), results for this grade level improved in the present study. Another conclusion is that in our sample, the ages at which Noelting assumed that students achieved each of the reasoning levels was delayed for Levels IIA and posterior levels, and consequently, students did not reach the upper stages of development assumed by Piaget and Inhelder (1951) for comparison of probabilities problems at the expected age.

When comparing the strategies used by our students with those from Pérez-Echeverría et al. (1986), they found that 10% of their students used multiplicative strategies for problems of comparing probabilities. These strategies include correspondence and proportionality that in our sample were used scarcely. There are, however, some exceptions for some problems, mainly for the strategies used by the Spanish students in Grades 8, 9, and 10. In summary, few students even in the upper grades in both countries used multiplicative strategies, which is why they failed to solve correctly the more difficult problems.

There was scarce use of incorrect strategies per item in the sample, although, when adding all the items, many students tried incorrect procedures to solve the harder problems. The incorrect strategies coincided with those described by Cañizares et al. (1997) and Green (1982). The proportion of students with biases such as equiprobability (Lecoutre & Durand, 1988) and the outcome approach (Konold, 1989), was very small. We observed more difficulty with probability comparison problems from the lower percentage of students at the higher reasoning levels, coinciding with the findings of Pérez-Echeverría et al. (1986) and results reported by Supply et al. (2020) with German students in the PISA examination.

Progress in developing proportional reasoning takes time and is completed by the transition from concrete to formal operations development (Lamon, 2007). This development seems to be slower in our sample and supports findings by Van Dooren et al. (2018) because many students still had difficulty with proportional reasoning after instruction.

The implications for teaching is that teachers should take into account these findings when teaching probability because when the proportional level of the item is low (for example, Levels IA, IB, and IIA), the students in all grades can easily solve the comparison of probabilities items. We recommend restricting the problems to these levels for students in these grades and progressively increasing the proportional reasoning level required by tasks in higher grades. In those grades, teachers should help students achieve higher levels of proportional reasoning, which is not only needed to solve probability problems but also needed in other areas such as geometry or algebra (Ben-Chaim et al., 2012).

Moreover, although proportional reasoning is no doubt a requisite for solving probability problems, competence in probability also influences the development of proportional reasoning. For this reason, teachers should take advantage of the different abilities that can be involved in comparing probabilities and use tasks related to this research to develop their students' mathematical ideas. These considerations should be included in the education of teachers to teach probability and proportionality. Finally, given the limitations of this study, more research on the relationships between proportional and probabilistic reasoning is needed, as suggested by Batanero and Álvarez-Arroyo (2023).

ACKNOWLEDGEMENTS

Grants PID2019-105601GB-I00 and PID2022-139748NB-100 funded by MCIN/AEI/10.13039/501100011033 and Fondo Europeo de Desarrollo Regional (FEDER, Una manera de hacer Europa).

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. https://www.apa.org/science/programs/testing/standards

Batanero, C., & Álvarez-Arroyo, R. (2023). Teaching and learning of probability. *ZDM Mathematics Education.* https://doi.org/10.1007/s11858-023-01511-5

Ben-Chaim, D., Keret, Y., & Ilany, B. S. (2012). *Ratio and proportion: Research and teaching in mathematics teachers' education*. Sense Publishers. https://doi.org/10.1007/978-94-6091-784-4_2

Berrocal, P. F. (1990). Relaciones teórico-empíricas entre los esquemas de proporción. Probabilidad y covariación [Theoretical-empirical relationships between ratio schemes: Probability and covariation]. *Revista de Psicología General y Aplicada, 43*(3), 331–337.

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: A literature review*. Nuffield Foundation.

Burgos, M., & Godino, J. D. (2020). Modelo ontosemiótico de referencia de la proporcionalidad. Implicaciones para la planificación curricular en primaria y secundaria [Ontosemiotic reference model of proportionality. Implications for curriculum planning in primary and secondary school]. *Avances de Investigación en Educación Matemática, 18,* 1–20. https://doi.org/10.35763/aiem.v0i18.255

Cañizares, M. J., Batanero, C., Serrano, L., & Ortiz. J. J. (1997). Subjective elements in children's comparison of probabilities. In E. Pehkonen (Ed.), *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education, Helsinki* (Vol. 2, pp. 49–56).

Carpenter, T. P., Fennema, E., & Romberg, T. A. (Eds.). (2012). *Rational numbers: An integration of research.* Routledge.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20(*1), 37–46.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Falk, R., Falk, R., & Levin, I. (1980). A potential for learning probability in young children. *Educational Studies in Mathematics, 11*, 181–204. https://doi.org/10.1007/BF00304355

Gal, I. (2005). Towards "probability literacy" for all citizens. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 43–71). Springer. https://doi.org/10.1007/0-387-24530-8_3

Gal, I., & Geiger. V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. *Educational Studies in Mathematics, 111*, 5–28. https://doi.org/10.1007/s10649-022-10151-7

Green, D. R. (1982). *Probability concepts in school pupils aged 11–16 years.* [Unpublished doctoral dissertation, University of Loughborough].

Hernández-Solís, L., Batanero, C., Gea, M. M., & Álvarez-Arroyo, R. (2021). Comparación de probabilidades en urnas: Un estudio con niños costarricenses de educación primaria [Comparison of probabilities at the ballot box: A study with Costa Rican elementary school children]. *Uniciencia, 35*(2), 1–18. https://doi.org/10.15359/ru.35-2.9

Hernández-Solís, L., Gea, M. M., Batanero, C., & Álvarez-Arroyo, R. (2023). Investigación sobre el razonamiento de los niños en la comparación de probabilidades [Research on children's reasoning in probability comparison]. *Boletín de Estadística e Investigación Operativa, 39*(1), 5–24.

Karplus, R., Pulos, S., & Stage, E. K. (1983). Early adolescents' proportional reasoning on "rate" problems. *Educational Studies in Mathematics, 14(*3), 219–233. https://doi.org/10.1007/BF00410539

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6,* 59–98. https://doi.org/10.1207/s1532690xci0601_3

Lamon, S. J. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 629–667). Information Age.

Lecoutre, M. P., & Durand, J. L. (1988). Judgements probabilistes et modèles cognitifs: Etude d'une situation aleatoire [Probabilistic judgments and cognitive models: Study of a random situation]. *Educational Studies in Mathematics, 19,* 357–368. https://doi.org/10.1007/BF00312452

Ministerio de Educación Pública. (2012). *Programas de Estudio de Matemáticas. I, II y III Ciclos de la Educación General Básica y Ciclo Diversificado* [Mathematics programs of Study I, II, and III. Cycles of general basic education and diversified cycle].

Ministerio de Educación, Cultura, y Deporte. (2014). *Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria* [Royal Decree 126/2014, of February 28, Establishing the basic curriculum for primary education].

Ministerio de Educación, Cultura, y Deporte. (2015). *Real decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la educación secundaria obligatoria y del bachillerato* [Royal Decree 1105/2014, of December 26, Establishing the basic curriculum for compulsory secondary education and the Baccalaureate].

Ministerio de Educación y Formación Profesional. (2022a). *Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria* [Royal Decree 157/2022, of March 1, Establishing the organization and minimum teaching requirements for primary education].

Ministerio de Educación y Formación Profesional. (2022b). *Real Decreto 217/2022, de 29 de marzo, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria* [Royal Decree 217/2022, of March 29, Establishing the organization and minimum education requirements for compulsory secondary education].

Muñiz-Rodríguez, L., Rodríguez-Muñiz, L. J., & Alsina, Á. (2020). Deficits in the statistical and probabilistic literacy of citizens: Effects in a world in crisis. *Mathematics, 8*(11), 1872. https://doi.org/10.3390/math8111872

Nikiforidou, Z. (2018). Probabilistic thinking and young children: Theory and pedagogy. In A. Leavy, M. Meletiou-Mavrotheris, & E. Paparistodemou (Eds.), *Statistics in early childhood and primary education* (pp. 21–34). Springer. https://doi.org/10.1007/978-981-13-1044-7_3

Noelting, G. (1980a). The development of proportional reasoning and the ratio concept. Part I. Differentiation of stages. *Educational Studies in Mathematics, 11*(2), 217–253.

Noelting, G. (1980b). The development of proportional reasoning and the ratio concept. Part II. Problem structure at successive stages: Problem solving strategies and the mechanism of adaptive restructuring. *Educational Studies in Mathematics, 11*(3), 331–363.

Pérez-Echeverría, M. P., Carretero, M., & Pozo, J. I. (1986). Los adolescentes ante las matemáticas: Proporción y probabilidad [Adolescents and mathematics: Proportion and probability]. *Cuadernos de Pedagogía, 133,* 9–13.

Piaget, J., & Inhelder, B. (1951). *La genése de l'idée de hasard chez l'enfant* [The genesis of the idea of chance in children]. Presses Universitaires de France.

Pratt, D., & Kazak, S. (2018). Research on uncertainty. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 193–227). Springer. https://doi.org/10.1007/978-3-319-66195-7_6

Resnick, L. B., & Singer, J. A. (1993). Proto quantitative origins of ratio reasoning. In T. P. Carpenter, E. Fennema & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 107–130). Lawrence Erlbaum.

Supply, A. S., Van Dooren, W., Lem. S., & Onghena, P. (2020). Assessing young children's ability to compare probabilities. *Educational Studies in Mathematics, 103*(1), 27–42. https://doi.org/10.1007/s10649-019-09917-3

Tourniaire, F., & Pulos, S. (1985). Proportional reasoning: A review of the literature. *Educational Studies in Mathematics, 16*(2), 181–204. https://doi.org/10.1007/PL00020739

Van Dooren, W., Vamvakoussi, X., Verschaffel, L., Marope, M., Vosniadou, S., Anderson, L., de Ibarrola, M., & Popa, S. (2018). *Proportional reasoning.* International Academy of Education.

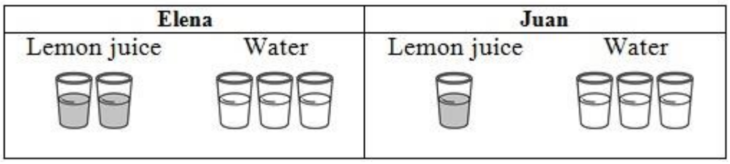
CARMEN BATANERO
Profesor Motos Guirao 4, 1, 180O2 Granada, Spain

APPENDIX

QUESTIONNNAIRE

FORM A

Item 1. Elena and Juan make some lemonade. Elena mixes 2 glasses of lemon juice with 3 glasses of water. Juan combines 1 glass of lemon juice with 3 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

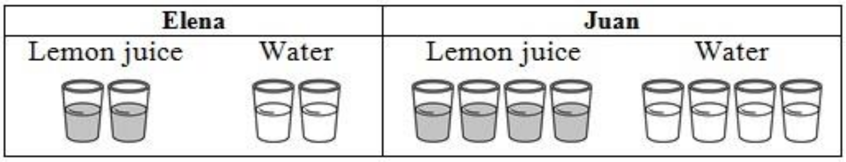


Which of the lemonades tastes more like lemon?

A ( ) Elena's.
B ( ) Juan's.
C ( ) Both are identical.
D ( ) I don't know.

Explain why you give this answer.

Item 3. Elena and Juan make some lemonade. Elena mixes 2 glasses of lemon juice with 2 glasses of water. Juan combines 4 glasses of lemon juice with 4 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

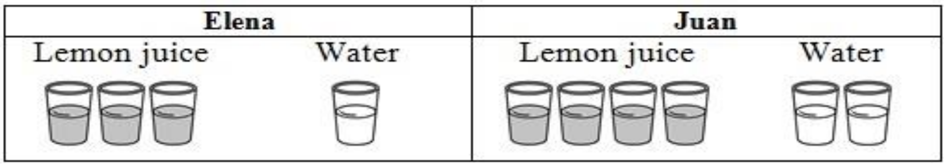


Which of the lemonades tastes more like lemon?

A ( ) Elena's.
B ( ) Juan's.
C ( ) Both are identical.
D ( ) I don't know.

Explain why you give this answer.

Item 5. Elena and Juan make some lemonade. Elena mixes 3 glasses of lemon juice with 1 glass of water. Juan combines 4 glasses of lemon juice with 2 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

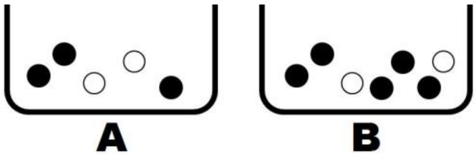


Which of the lemonades tastes more like lemon?

A ( ) Elena's.
B ( ) Juan's.
C ( ) Both are identical.
D ( ) I don't know.

Explain why you give this answer.

Item 7. There are 3 black counters and 2 white counters in box A. In box B, there are 5 black counters and 2 white counters. Look at the picture.

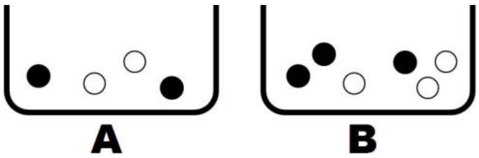


We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.

Item 9. There are 2 black counters and 2 white counters in box A. In box B, there are 3 black counters and 3 white counters. Look at the picture.

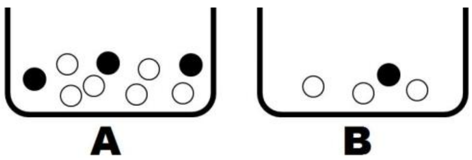


We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.

Item 11. There are 3 black counters and 6 white counters in box A. In box B, there are 1 black counters and 3 white counters. Look at the picture.



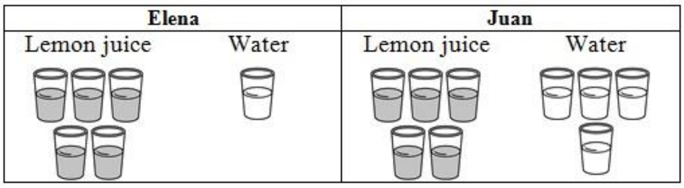
We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.

FORM B

Item 2. Elena and Juan make some lemonade. Elena mixes 5 glasses of lemon juice with 1 glass of water. Juan combines 5 glasses of lemon juice with 4 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

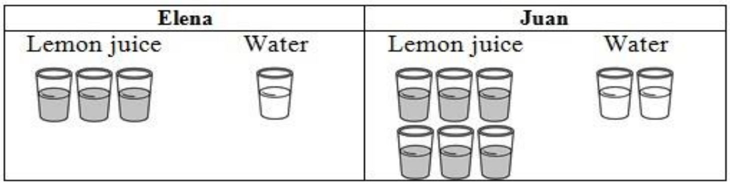


Which of the lemonades tastes more like lemon?

A ( ) Elena’s.
B ( ) Juan’s.
C ( ) Both are identical.
D ( ) I don’t know.

Explain why you give this answer.

Item 4. Elena and Juan make some lemonade. Elena mixes 3 glasses of lemon juice with 1 glass of water. Juan combines 6 glasses of lemon juice with 2 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

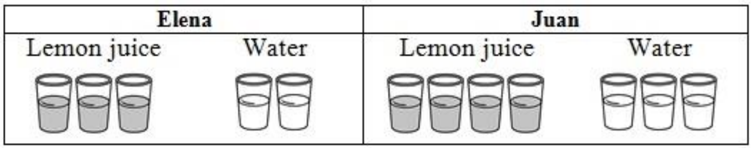


Which of the lemonades tastes more like lemon?

A ( ) Elena’s.
B ( ) Juan’s.
C ( ) Both are identical.
D ( ) I don’t know.

Explain why you give this answer.

Item 6. Elena and Juan make some lemonade. Elena mixes 3 glasses of lemon juice with 2 glasses of water. Juan combines 4 glasses of lemon juice with 3 glasses of water. All glasses contain the same amount of liquid. Look at the picture.

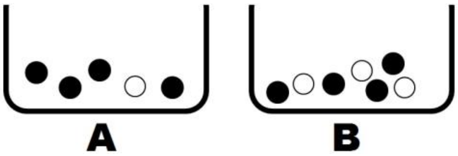


Which of the lemonades tastes more like lemon?

A ( ) Elena’s.
B ( ) Juan’s.
C ( ) Both are identical.
D ( ) I don’t know.

Explain why you give this answer.

Item 8. There are 4 black counters and 1 white counter in box A. In box B, there are 4 black counters and 3 white counters. Look at the picture.

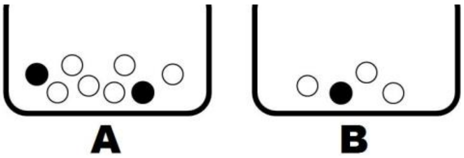


We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.

Item 10. There are 2 black counters and 6 white counters in box A. In box B, there is 1 black counter and 3 white counters. Look at the picture.

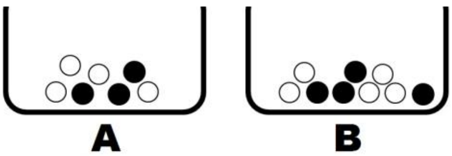


We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.

Item 12. There are 3 black counters and 4 white counters in box A. In box B, there are 4 black counters and 5 white counters. Look at the picture.



We shake the boxes and pick a counter with our eyes closed. In which box are we more likely to pick a black counter? Mark the correct answer:

A ( ) You are more likely to draw a black counter from box A.
B ( ) You are more likely to draw a black counter from box B.
C ( ) There is an equal chance of drawing a black counter in both boxes.
D ( ) I don't know.

Explain why you give this answer.