Journal of Mathematics Education
December 2009, Vol. 2, No. 2, pp.52-67

A Comparative Study on Junior High School Students' Proof Conceptions in Algebra between Taiwan and the UK

Erhtsung Chin
National Changhua University of Education, Taiwan
Foulai Lin
National Taiwan Normal University, Taiwan

This paper aims to investigate Taiwanese junior high school students' proof conceptions in algebra with an incidental purpose to compare the results with the study of Healy and Hoyles (2000). Questionnaire survey was adopted as the main research method in the study. The 1059 surveyed subjects aged 14-15 years are nationally random-sampled by means of a two-stage sampling. The students' proof conceptions are focused on: the arguments they would adopt for their own approaches; the arguments they consider would receive the best mark; assessing the validity and explanatory power of the given arguments; and the competence of constructing a proof. The main results of students' responses to two multiple-choice questions and their constructed proofs are analyzed and discussed. In addition, 136 teachers also responded to some selected items for providing data to analyze the teachers' factor.

Key words: algebra, comparative study, proof conception.

Introduction

The issue of getting students to understand the role of reasoning and proving in mathematics has been drawing much attention within the community of mathematics education for decades. Many research studies indicate that students encounter difficulties when making mathematical proofs (Balacheff, 1990; Chazan, 1993; Duval, 1998; Harel & Sowder, 1998; Moore, 1994; Tall et al., 2001). Some mathematics educators also try to grasp students' transition from informal to formal reasoning (Hoyles & Healy, 1999; Moore, 1994). However, some different voices appear in North America. For over the past several decades, many mathematics educators have suggested that proof should be degraded to a lesser role in the secondary mathematics curriculum. But some researchers hold an opposite view by retorting that none of the factors they offered justifies such a move and asserting the value of proof in the classroom "both as a reflection of its central role in mathematical practice and as an important tool for the promotion of understanding" (Hanna, 1997, p.171).

There exist many factors influencing students' learning mathematical proof. Based on the construct of "proof schemes" (Harel & Sowder, 1998), Harel and Sowder (2007) generalize a comprehensive perspective on the teaching and learning of proofs which consists of a range of factors including: mathematical and historical-epistemological, cognitive, and instructional-socio-cultural. Most of the empirical studies focus mainly on the cognitive factor by analyzing students' answers to questions requiring proof. In addition, students' conceptions about proof that also influence their approaches to argumentation and proof (Healy & Hoyles, 1998; 2000; Simon, 1996) are categorized in the cognitive factor (Harel & Sowder, 2007).

In the little empirical evidence offered in the corpus of mathematics education research to document students' views of proof and the relationship between their views and approaches to proof, Healy and Hoyles conducted a multi-year national research project entitled "Justifying and Proving in School Mathematics in England and Wales" (Healy & Hoyles, 1998), which included a survey of high-attaining 14- and 15-year-old students about proof in algebra (Healy & Hoyles, 2000). In Taiwan, our students' performances in mathematics and science have correctly been evaluated as leading in the world (e.g. PISA 2006; TIMSS 2003). Nevertheless, as mathematics educators, we are interested in their competence of mathematical argumentation and reasoning, especially because the algebraic proofs have been removed from our national curriculum. Therefore, as referring to Healy and Hoyles' survey study, we also organized a national research project aimed at investigating Taiwan junior high school students' competence of mathematical argumentation and reasoning, and formulating their learning trajectories when making mathematics argumentation. It has to be noted that the subjects of Healy and Hoyles' study were high-attainers who were at the top 20-25% of the student population, but the subjects in Taiwan were nationally random-sampled.

The term "proof conception" (or conception of proof) had not been really defined by researchers when it appeared in academic articles (e.g. Harel & Sowder's studies, Hoyles and her colleagues' studies). In 2007, in her PhD dissertation about student teachers' conceptions of proof and facilitation of argumentation in secondary mathematics classrooms, Conner gives a definition of proof conception by defining "a person's conception of proof as the person's ability to prove and analyze arguments, perception of the role and need for proof in mathematics, and affective perception of proof" (Conner, 2007). In our study, since the subjects are junior high school students who just starting to encounter what is called "mathematical proof," it seems unnecessary to assess their perceptions of the role and need for proof and affective perceptions of proof. Therefore, as referring to Healy and Hoyles (1998) and other researchers' works in various aspects of argumentation and proof, we decided to focus the students' proof conceptions on: (1) the arguments they would adopt for their own approaches; (2) the arguments they consider would receive the best mark; (3) assessing the validity of the given arguments; (4) assessing the explanatory power of the given arguments; and (5) the competence of constructing a proof. These five elements should be concrete and sufficient for investigating the students' abilities to prove and analyze arguments.

The comparative research in mathematics education between the East and the West has been a hot issue for decades. There exists great variation in mathematics classroom culture between Taiwan and the UK (e.g. Lin, 1988). At the end of Healy and Hoyles' paper (2000), they seem to suggest some international comparisons for possible future studies. Consequently, an incidental purpose of the study is to compare our results with Healy and Hoyles' in order to shed light on any valuable findings.

Method

The Research Design

This paper reports some of the main results of a three-year national research project, "Taiwanese Junior High School Students' Competence of Mathematical Argumentation and Reasoning." This project was aimed at investigating students' conceptions of proofs and the characteristics of arguments recognized as proofs by the students, and formulating the students' learning trajectories when making mathematics argumentations in Taiwan. This project was conducted by a research team which consisted of six mathematics educators with doctorates in math education, one mathematician with a doctorate in mathematics, two senior high school math teachers with about twenty years teaching experience, and one Ph.D. and two M.S. research students in math education. The two domains of proof explored in the study are arithmetic/algebra and geometry. Questionnaire survey method was adopted as the main research method in the study. The content of the two questionnaires consisted of the two domains while each domain with respect to grades 7, 8 and 9 (aged 12-15 years) were designed. The two questionnaires, each containing three booklets, were piloted for confirming the content validity, and then were formally administered by means of the two-stage sampling approach. The students' responses were coded and descriptive statistics was applied in the analysis. In addition, teachers' responses to some selected items were also collected to analyze the teachers' factor. Detailed research methods are elaborated in the as following graphs.

Instrument

Based on the reviewed literature, and referring to the proof questionnaire designed by Hoyles and her colleagues (e.g. Healy & Hoyles, 1998; 2000; Hoyles & Kuchemann, 2002), the relevant constructs of the two proof domains arithmetic/algebra and geometry were listed. In addition, the research team also tried to collect and devise the problem items. Then the research team was divided into two groups, each with six members. Each group took charge of developing a two-way specification table of relative constructs and the corresponding problem items with respect to the domain of arithmetic/algebra or geometry. Afterwards, the two groups exchanged their tasks and did the categorizations again. The whole group would discuss in order to reach agreement when there was inconsistency of the categorization. The first draft of the questionnaires was piloted with 60 students to assess whether they could understand all the items and could answer it within the given 45 minutes. Modifications were made based on their responses. The second draft of the questionnaire was piloted again with a total of 1534 seventh-ninth graders in two schools to confirm the validity of the second modification of the questionnaire.

The content of items in the questionnaire included views and conceptions about different proof approaches, true or false statements, specialization and application of a proved statement or a specific argumentation, local reasoning, constructing proofs, validity of a conditional statement, cognition of number patterns, and cognition of representation translating. In this article, we only report some findings from the booklet of grade 9 of the domain of arithmetic/algebra. The related algebraic items designed to probe students' conceptions of proof included two types. First, students were presented a mathematical conjecture and a range of purported arguments in support of them;

they were asked to make two selections from these arguments — the argument that would be closest to their own approaches and the one which they considered would receive the best mark from their teachers. Second, students were also asked to make assessments of these given arguments in terms of their validity and explanatory power. There were three evaluations of the validity of each argument (mistake, sometimes true, always true), as well as two evaluations of explanatory power (shows you why, explains to someone in your class). Two conjectures offered to students included one familiar (A1) and another unfamiliar (A2) (see Appendix). These two conjectures first appeared in the UK's version (see Healy & Hoyles, 2000, pp. 400-401), but, based on the practice of our Taiwanese students, we modified some arguments in the familiar conjecture by deleting a visual argument which is not a possible answer for our students, and adding another three wrong algebraic arguments with single variable which are the typically wrong styles of answers made by Taiwanese students. Besides, two items, one familiar and another unfamiliar, with an open format to ask students to construct their own proofs were included in our analysis in the article. The two questions are as follows:

(A3) Prove that when you add any 2 odd numbers, your answer is always even;

(A4) Prove that if p and q are any two odd numbers, $(p+q)\times(p-q)$ is always a multiple of 4.

Participants

The surveyed subjects were nationally random-sampled by means of the two-stage sampling approach, different from the UK's subjects who were high-attainers (the top 20-25% of the student population). The first stage of the sampling was to divide Taiwan into six areas, then to calculate the number of students to pick from each area based on the ratio of population of the junior high school students in each area to the whole country, and finally to decide the number of schools to pick in each area in accordance with the average number of students in each school within that area (the probability of a school being picked is the ratio of the student number of that school to the area to which the school belongs). The second stage of sampling was to number the classes of each grade in the picked school from 1-13, then to pick 2 from the 13 groups of classes of each grade (since this is an intergraded project of 12 sub-projects, the rest groups of students were surveyed by the other sub-projects). Half of each picked class of students answered the booklet of arithmetic/algebra, and the other half geometry. The numbers of subjects answering different booklets of the questionnaires were within a range from 1059 to 1181. In this paper, only the 1059 9^{th} graders who answered the booklet in arithmetic/algebra were examined. The percentages of valid responses to the items A1-A4 were within 91.0% to 92.6%. In addition, 136 teachers also responded to the revised items A1 and A2 (see Appendix), providing data to analyze the teachers' factor, while the percentage of valid responses was 100%.

Data Analysis

Students' and teachers' responses to A1 and A2, as well as students' evaluations of the validity and explanatory power of each given argument were coded (for "validity rating" (mistake, sometimes true, always true), 2=correct evaluation, 1=partially correct evaluation, 0=incorrect; for "explanatory power", 2=explains private and public, 1=explains private or public, 0=does not explain). Students' constructed proofs of A3 and A4 were scored from 0 to 3, in which "0" means "no basis for the construction of a correct proof", "1" means "no deductions but relevant information presented", "2" means "partial proof including all information needed but some reasoning omitted", and "3" means "complete proof", as well as categorized by the forms of proof as "empirical", "narrative", "formal (algebraic)", "none (blank)", and "other", in which "empirical" means "only producing empirical(numerical) example(s)", "narrative" means "giving an informal argument in a narrative style without using algebraic representation", and "formal (algebraic)" means "offering a formal argument with algebraic representation". The coding and marking was conducted by 29 undergraduate students who majored in math while each response was reviewed by two reviewers. If there was inconsistency between the two reviewers' results, the research team would examine it again and make the final decision. Descriptive statistics based on frequency tables, simple correlations, and tests of significance were produced by means of SPSS 13.0.

Results and Discussion

Students' performances in the four items A1–A4 and teachers' responses to A1 and A2 will be analyzed and interpreted mainly by means of descriptive statistics. In addition, the main results are also compared with the UK's results (Healy & Hoyles, 2000). Please note that the item A1 in Taiwan's version corresponds to A1 in the UK's version, A2 (Taiwan) to A6 (UK), A3 (Taiwan) to A4 (UK), and A4 (Taiwan) to A7 (UK).

Students' and Teachers' Responses to Items A1 and A2

Table 1 presents the distribution of Taiwanese and UK students' and teachers' choices in the multiple-choice questions A1 and A2.

Table 1

Distributions of Taiwan (TWN) and UK Students' and Teachers' Choices of Proofs for A1 and A2 (TWN/UK)

	Percentages of students		Percentages of teachers	
Argument	Own approach	Best mark	Own approach	Best mark
Argument chosen for A1	TWN/UK *N*=981/2450	TWN/UK *N*=964/ 2423	TWN/UK *N*=136/94	TWN/UK *N*=136/94
Arthur (algebraic)	16/12	18/22	84/81	49/62
Bonnie (empirical)	28/24	8/3	1/3	7/7
Ceri (narrative)	4/17	8/18	6/10	7/11
Duncan (narrative)	19/29	11/7	3/6	21/12
Eric (algebraic)	5/2	13/42	0/0	1/9
Yvonne (visual)	---/16	---/9	---/---	---/---
Ian (algebraic)	9/---	14/---	3/---	7/---
Howard (algebraic)	13/---	19/---	3/---	6/---
Peggy (algebraic)	6/---	9/---	0/---	2/---
Argument chosen for A2	TWN/UK *N*=980/2381	TWN/UK *N*=971/2348	TWN/UK *N*=136/94	TWN/UK *N*=136/94
Kate (narrative)	31/41	16/19	68/70	29/48
Leon (empirical)	31/39	7/2	2/4	7/7
Maria (algebraic)	19/13	19/24	4/3	7/6
Nisha (algebraic)	19/7	58/55	26/23	57/39

Note. "---" in the table means the option did not appear in the booklet.

Without respect to A1 or A2, the differences between the choices Taiwanese students made for their own approaches and for best mark appeared highly significant (A1: χ^2=198.5, *df*=7, $p < .001$; A2: χ^2=227.2, *df*=3, $p < .001$). In the familiar conjecture A1, as to students' own approaches, 49% of students choose algebraic arguments while 33% picked the wrong ones. Twenty-eight percent of students chose the only empirical argument, which is wrong. The rest 23% chose the two narrative arguments, which are both correct. As to students' choices for best mark, 73% choose the algebraic while 55% picked the wrong choices. It decreases to only 8% choosing the empirical. In the unfamiliar conjecture A2, as to students' own approaches, 31% chose both narrative and empirical, while 19% selected both the other two algebraic. As to students' choices for best mark, the most complex algebraic one, which is correct, was most favored (58%), while the empirical one was chosen only 7% of the time.

Compared with the UK's results, their high-attainers' responses to A1 and A2 both demonstrated highly significant differences (A1: χ^2=1741.5, *df*=5, $p<.0001$; A2: χ^2=1891.2, *df*=3, $p<.0001$, Healy & Hoyles, 2000, p.407). From the differences of the χ^2 and *p* values between Taiwan's and the UK's results (the UK's χ^2 values are much higher than Taiwan's while its *p* values are lesser), it seems to show more significant difference in the UK than in Taiwan even though the number of UK's subjects was more that Taiwan's. It can be obviously noticed that the arguments that were the most popular for the UK's students' own approaches turned out to be the least popular for best mark, and vice versa. In the UK's students' responses, narrative (Duncan, which is correct) and empirical (Bonnie, incorrect) arguments were popular for one's own approach but not for best mark. This phenomenon seems not to occur with Taiwanese students since the preferences of Taiwanese students were more equalized. However, algebraic arguments appeared to be popular for their choice for best mark (Eric, incorrect, and Arthur, correct), which was consistent with the UK's students.

Similar to the students' responses, no matter in A1 or A2, the differences between Taiwanese teachers' choices were highly significant (A1: χ^2=43.1, *df*=7, $p<.001$; A2: χ^2=42.5, *df*=3, $p<.001$). About the teachers' responses to A1, it is not surprising that 84% of the teachers chose the correct algebraic argument (Arthur) for their own approach, while only 49% predicted the students would select it for best mark. In addition, 21% of teachers predicted the students would select the correct narrative argument (Duncan) for best mark. In A2, as to teachers' choices for their own approaches, 68% chose the narrative (Kate), which is a correct argument, and 26% picked the complex correct algebraic one (Nisha). As to the argument, 57% chose the complex algebraic one (Nisha), while 29% picked the narrative (Kate).

From the above results, Taiwanese students seemed to prefer the algebraic arguments and refuse the empirical for best mark. However, the teachers' views appeared to be a bit different. They seemed to consider the beauty of logic as a crucial factor rather than simply form. About students' evaluations of the validity and explanatory power of each given argument (Table 2), Taiwanese students were much more likely to correctly identify a wrong algebraic proof with nonsense symbols (e.g. Eric) than a correct algebraic argument with concise reasoning (e.g. Arthur). They might not correctly evaluate the generality of narrative and algebraic arguments (e.g.

Duncan & Arthur) whereas the UK's high-attainers performed much better on the narrative argument, even though they thought these arguments had explanatory power. They appeared to have clear ideas of the validity of the two empirical arguments (Bonnie & Leon), but they also considered these empirical arguments could show them that the given conjectures were correct. The reason should be that the students considered these empirical arguments could help them make sure of the correctness of the conjectures, but would not be sufficient for offering valid proofs for these conjectures.

Table 2

Students' Evaluations of Validity Ratings and Explanatory Power of Empirical, Narrative and Algebraic Arguments (in percentages, TWN/UK)

	Validity ratings			Explanatory power		
Argument type	0	1	2	0	1	2
Empirical arguments						
Bonnie (A1)	9/37	10/9	80/54	22/24	28/51	50/25
Leon (A2)	9/28	15/12	76/60	17/33	28/48	45/19
Narrative argument						
Duncan (A1, correct)	52/24	15/6	33/68	29/18	23/40	48/42
Algebraic arguments						
Arthur (A1, correct)	50/44	18/15	32/40	30/56	33/33	37/11
Eric (A1, incorrect)	17/69	18/19	65/12	63/64	23/33	14/3

Students' Constructed Proofs of A3 and A4

Table 3 shows the distribution of forms of presentation for students' constructed proofs of A3 and A4. In the familiar conjecture A3, 42% of Taiwanese students offered empirical arguments, and 32% provided algebraic arguments, whereas only 8% produced narrative arguments. As to the unfamiliar conjecture A4, the percentage of students offering empirical arguments increased to 50%, and the percentage providing algebraic arguments decreased to 11%, whereas the narrative increased to 13%. Compared with the UK's high-attainers, giving numerical examples (e.g. empirical arguments) was the most popular approach when constructing their own proofs both in Taiwan and the UK. It is more significant when facing an unfamiliar conjecture. A rather divergent phenomenon is that "narrative" was a more popular form in the UK than in Taiwan, whereas more Taiwanese students tried to offer algebraic arguments than did the British. This phenomenon is roughly consistent with the distribution of students' choices of their own approaches. There is difference between Taiwanese students or the UK's high-attainers, although most of them were aware that empirical arguments had limitations, they could only produce some examples to evaluate the given conjectures without being able to offer reasonable argumentations. This phenomenon was even more obvious with a harder conjecture.

Table 3

Distribution of Forms of Presentation for Constructed Proofs of A3 and A4 (TWN/UK)

	Familiar conjecture (A3)		Unfamiliar conjecture (A4)	
Form of proof	No.	%	No.	%
Empirical	451/845	42/34	525/1062	50/43
Narrative	83/692	8/28	137/792	13/32
Formal (algebraic)	334/281	32/11	121/82	11/3
None	183/74	17/3	266/443	25/18
Other	8/567	1/24*	10/80	1/4

Note. *In the UK's result, 8% of the responses for A4 were attempts at visual proofs, and 15% were attempts to produce an exhaustive proof by examples referring to the unit digit.

As shown in Table 4, 37% of Taiwanese students were classified as "partial" or "complete proof" in A3, and 20% in A4. Compared with the UK's high-attainers, Taiwanese students performed much better in the unfamiliar conjecture A4, and about the same in the familiar conjecture A3. This result seems to reflect the fact that Taiwan has been highly ranked in recent international assessments, e.g. PISA 2006, TIMSS 2007 (Mullis et al., 2008) & TIMSS 2003 (Mullis, et al, 2004). It should be noted that in Taiwan the current curriculum for junior high school mathematics is algorithm-oriented, in which the formal algebraic proof has been removed. Students are only taught to verify some algebraic conjectures but not asked to prove them formally.

Table 4

Distribution of Scores for Constructed Proofs of A3 and A4 (TWN/UK)

	Familiar conjecture (A3)		Unfamiliar conjecture (A4)	
Constructed proof score	No.	%	No.	%
0 No basis for the construction of correct proof	218/354	20/14	396/866	37/35

1 No deductions but relevant information presented	451/1130	43/46	455/1356	43/55
2 Partial proof, including all information needed but some reasoning omitted	330/438	31/18	145/154	14/6
3 Complete proof	60/537	6/22	63/83	6/3

We also compared the total number of Taiwanese students who chose a correct argument in A1 and A2 with the total number of students who construct either a partial or complete proof for A3 and A4 for checking the difference. Only the χ^2 test between A1 and A3 (which were both familiar conjectures) did not reach the significance level (χ^2=.459, df=1, p=.498 > .05). The other three all showed highly significant (A1vs A4, χ^2=75.043, df=1, p < .001; A2 vs A3, χ^2=27.143, df=1, p < .001; A2 vs A4, χ^2=170.041, df=1, p < .001) which might suggest that the students were significantly better at selecting correct arguments than at constructing them on their own. However, the case in the UK is more obvious since the χ^2 value is much higher and the p value is less than .0001 (χ^2=1088.77, p < .0001).

Conclusion and Implication

Even though the study showed that our Taiwanese students performed better than the UK's high-attainers, most of them were still unable to construct satisfactory proofs on their own. Although the majority of Taiwanese students would select algebraic arguments for their own approaches and were able to value the validity of empirical arguments, nearly half of the students produced empirical arguments when making their own proofs. A fairly discrepant phenomenon is that narrative arguments were popular for the UK's students' constructed proofs and in the selections of their own approaches, but not for our Taiwanese students. In addition, most of Taiwanese students even failed to evaluate the validity of the given narrative conjectures (e.g. Duncan in table 2). The reason might be related to the teaching most Taiwanese students received, which was lecture-oriented but not cooperative learning oriented. Therefore, students were short of experiences in sharing ideas with peers or the teacher, which made them unfamiliar with using the narrative approach to make argumentations.

Having the chance to conduct this research made us able to gain a deeper insight of our students' conceptions and competence of proofs and proving in algebra. Based on the findings of the study, we could propose some teaching strategies for helping improve the students' proving competence, even though Taiwanese students' mathematics performance was ranked first in PISA 2006 and TIMSS 2007 recently. "Empirical arguments" were evaluated by the students with high explanatory power but low validity, however, they still predominated most students' constructed proofs. Actually, checking examples plays an important role in problem solving or mathematical inquiry. Through checking examples, a pattern should be spotted, and elaborated, or proved when solving a mathematical problem. Therefore, as our students predominantly use empirical arguments for their own proofs, our teaching could focus on guiding the students to practice generalizing and specializing, i.e. induction and deduction reasoning. Besides, encouraging the students to communicate with others in the class would be also important for improving their argumentation and proving competence.

Acknowledgement

The study presented in this article is funded by the National Science Council of Taiwan with project number NSC97-2511-S-018-001 MY2.

References

Balacheff, N. (1990). A study of students' proving processes at junior high school level. In I. Wirszup & R. Streit (Eds.), *Developments in school mathematics education around the world* (pp. 284-297). Chicago: NCTM.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Research in Mathematics, 24,* 359-387.

Conner, A. (2007). *Student teachers' conceptions of proof and facilitation of argumentation in secondary mathematics classrooms.* Unpublished doctoral dissertation, The Pennsylvania State University, University Park.

Duval, R. (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.), *Perspectives on the teaching of geometry for 21st century* (pp. 37-52). Dordrecht Netherlands: Kluwer.

Hanna, G. (1997). The ongoing value of proof. *Journal Mathematik Didaktik, 18,* 171-185.

Harel, G. & Sowder, L. (1998). Students' proof schemes: results from exploratory studies. In A.H. Schoenfeld, J. Kaput & E. Dubinsky (Eds.), *Research in collegiate mathematics education*, (Vol. 3, pp.234-283). Providence, RI: American Mathematical Society

Harel, G. & Sowder, L. (2007). Toward the comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805-842). Charlotte, NC: Information Age.

Healy, L. & Hoyles, C. (1998). *Justifying and proving in school mathematics* (Technical Report). London: Institute of Education, University of London.

Healy, L. & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education, 31*, 396-428.

Hoyles, C., & Healy, L. (1999). Linking informal argumentation with formal proof through computer-integrated teaching experiment. In O. Zaslavsky (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education, 3*, 105-112. Israel: Haifa University.

Hoyles, C. & Kuchemann, D. (2002). Students' understanding of logical implication. *Educational Studies in Mathematics, 51*, 193-223.

Lin, F. L. (1988). Societal differences and their influences on children's mathematics understanding. *Educational Studies in Mathematics, 19*, 471-497.

Moore, R. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27*, 249-266.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M. O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Simon, M. (1996). Beyond inductive and deductive reasoning: The research for a sense of knowing. *Educational Studies in Mathematics, 30*, 197-210.

Tall, D. O., Gray, E. M., Ali, M. b., Crowley, L., Demarois, P., McGowen, M., Pitta, D., Pinto, M., Thomas, M., & Yusof, Y. b. M., (2001). Symbols and the Bifurcation between Procedural and Conceptual Thinking, *Canadian Journal of Science, Mathematics and Technology Education*, *1*, 81–104.

Appendix

A1. Some students were trying to prove whether the following statement is true or false: When you add any 2 even numbers, your answer is always even.

Arthur's answer:
a, b are any whole numbers,
$2a$ and $2b$ are any two even numbers
$2a+2b=2(a+b)$
So Arthur says it's true.

Bonnie's answer:
$2+2=4$ $4+2=6$ $2+4=6$
$4+4=8$ $2+6=8$ $4+6=10$
So Bonnie says it's true.

Ceri's answer:
Even numbers are numbers that can be divided by 2. When you add numbers with a common factor, 2 in this case, the answer will have the same common factor.
So Ceri says it's true.

Duncan's answer:
Even numbers end in 0, 2, 4, 6, or 8. When you add any two of these, the answer will still end in 0, 2, 4, 6, or 8.
So Duncan says it's true.

Yvonne's answer (only in UK's version):
●●●●● ●●●● ●●●●●●●●●
●●●●● + ●●●● = ●●●●●●●●●
So Yvonne says it's true.

Eric's answer:
Let x, y are any whole numbers,
$x+y=z$
$z-x=y$
$z-y=x$
$z+z-(x+y)=x+y=2z$
So Eric says it's true.

Howard's answer (only in Taiwan's version):
x is any odd number, then $x+1$, $x-1$ are even numbers, $(x+1)+(x-1)=2x$
So Howard says it's true.

Ian's answer (only in Taiwan's version):
Let a be any even number, then the next even number will be $a+2$.
$a+(a+2)=2a+2=2(a+1)$
So Ian says it's true.

Peggy's answer (only in Taiwan's version):
$2x$ is an even number,
$2x+2x=4x=2\times 2x$
So Peggy says it's true.

1. From the above answers, choose one that would be closest to what you would do if you were asked to answer this question.

2. From the above answers, choose the one to which your teacher would give the best mark. (In the teacher's version, A2 is changed as "*choose the one your students would select for best mark*")A2. Kate, Leon, Maria, and Nisha were asked to prove whether the following statement is true or false: When you multiply any 3 consecutive numbers, your answer is always a multiple of 6.

Kate's answer:

A multiple of 6 must have factors of 3 and 2. If you have three consecutive numbers, one will be a multiple of 3 as every third number is in the three times table. Also, at least one number will be even and all even numbers are multiples of 2. If you multiply the three consecutive numbers together, the answer must have at least one factor of 3 and one factor of 2.

So Kate says it's true.

Leon's answer:

$1 \times 2 \times 3 = 6$

$2 \times 3 \times 4 = 24$

$4 \times 5 \times 6 = 120$

$6 \times 7 \times 8 = 336$

So Leon says it's true.

Maria's answer:

x is any whole number

$x \times (x+1) \times (x+2) = (x^2+x) \times (x+2)$

$= x^3 + x^2 + 2x^2 + 2x$

Cancelling the xs gives $1+1+2+2=6$

So Maria says it's true.

Nisha's answer:

Of the three consecutive numbers, the first number is either EVEN, which can be written $2a$(a is any whole number), or ODD, which can be written $2b-1$(b is any whole number).

If EVEN

$2a \times (2a+1) \times (2a+2)$ is a multiple of 2

and either a is a multiple of 3 DONE

or a is not a multiple of 3, so $2a$ is not a multiple of 3

then either($2a+1$)is a multiple of 3 or($2a+2$)is a multiple of 3 DONE

If ODD

$(2b-1) \times 2b \times (2b+1)$ is a multiple of 2

and either b is a multiple of 3 DONE

or b is not a multiple of 3, so $2b$ is not a multiple of 3

then either ($2b-1$)is a multiple of 3 or($2b+1$) is a multiple of 3 DONE

So Nisha says it's true.

1. From the above answers, choose one that would be closest to what you would do if you were asked to answer this question.

2. From the above answers, choose the one to which your teacher would give the best mark. (In the teacher's version, A2 is changed as "*choose the one your students would select for best mark*")

Authors:
Erhtsung Chin
National Changhua University of Education, Taiwan
Email: abechin@cc.ncue.edu.tw

Foulai Lin
National Taiwan Normal University, Taiwan
Email: linfl@math.ntnu.edu.tw