European Journal of Science and Mathematics Education, 2024, **12**(1), 112-127
ISSN: 2301-251X (Online)

OPEN ACCESS

Research Article

Assessing the added value of a history-based activity for students with low mathematics skills

Thomas De Vittori [1*]
0000-0002-8792-3283

Gaëlle Louaked [2]
0009-0000-9574-2100

Marie-Pierre Visentin [3]
0009-0007-2851-0079

[1] Laboratoire de Mathématiques de Lens, Faculté des Sciences Jean Perrin, Université d'Artois, Arras, FRANCE
[2] Laboratoire Paul Painlevé, Université de Lille, Lille, FRANCE
[3] Primary School Henri-Matisse, Saint Sulpice, FRANCE
[*] Corresponding author: thomas.devittori@univ-lille.fr

Citation: De Vittori, T., Louaked, G., & Visentin, M.-P. (2024). Assessing the added value of a history-based activity for students with low mathematics skills. *European Journal of Science and Mathematics Education, 12*(1), 112-127. https://doi.org/10.30935/scimath/13868

ARTICLE INFO

Received: 07 Jul 2023

Accepted: 26 Oct 2023

ABSTRACT

The aim of this pilot study is to evaluate the relevance of the use of history in mathematics education. This paper presents an experiment carried out in France with sixth-grade students (n=108) in which an ancient number system is used, an approach that is commonly suggested in French sixth-grade textbooks but has previously been unassessed. Based on the data of a pre-test and a post-test surrounding an activity on an ancient Chinese numeration system, a statistical analysis using Rasch modeling shows a specific added value of the history of mathematics for students with low abilities in mathematics. For these students, a significant increase in observed abilities of +0.67 logit in mean is measured with a large effect size (Cliff delta +0.52). This effect is then weighted by considering the regression to the mean (RTM) effect, leading to a value around +0.14 logit in mean and a negligible effect size (Cliff delta +0.10). So, this pilot study shows the important effect of RTM, which suggests a very strong rebalancing of students' results. In the last part of the paper, we discuss how RTM can nonetheless be positively interpreted in this specific context where students' disorientation is one of the purposes of history in mathematics education.

Keywords: history of mathematics, number system, sixth grade, Rasch model, regression to the mean

INTRODUCTION

According to the introduction to a recent international ICMI study (Bartolini Bussi & Sun, 2018), whole numbers and the fundamental principles of arithmetic can be regarded as the cornerstone of mathematics education. Beginning in kindergarten, these concepts are introduced and gradually developed throughout elementary school, laying a strong foundation for the concepts and methods taught in secondary school and beyond. Nevertheless, even with the recognized importance of this foundational learning and the allocated classroom time, students frequently face difficulties in fully comprehending the number system. Numerous students demonstrate limited conceptualization of decimal numbers, as observed in the research by Chesné and Fischer (2015), and sometimes struggle with whole numbers. According to Houdement and Tempier (2019), the grasp of numbers in terms of units, tens, hundreds, and beyond serves as a pivotal factor for learning and serves as an indicator of students' proficiency. These numerical decompositions are present in verbal expressions but are not explicitly reflected in written form, creating a disconnect that present a true challenge for certain students. This challenge remains persistent throughout students' academic journey, often resulting in difficulties with numerical tasks (Chambris & Tempier, 2017) and calculations or problem-solving (Chesné & Fischer, 2015). Clearly, the key to our numerical system lies in the transition from one place value to another by consolidating ten representatives of a lower-number unit (e.g., 10) into a single representative of the higher-number unit (e.g., 100). This process, particularly in base 10, forms the foundation for computational algorithms, demanding proficiency in various place values. As highlighted by Houdement and Tempier (2019), early research identified a strong correlation between inadequate numerical comprehension and poor performance in computational tasks (Barr, 1978; Lambert & Moeller, 2019; Ma, 1999; Thanheiser, 2012). Specifically, the authors note variations in performance when transitioning from two-digit calculations to three-digit calculations, as highlighted in the work of Thomas (2004). It is within these calculations that deficiencies in handling larger numbers become evident, as indicated by research conducted by Baturo (2000) and Bednarz and Janvier (1992). Furthermore, the significance of classes, such as units, thousands, millions, and so on in which place values (units, tens, and hundreds) are reiterated, becomes major learning topic. In France, as highlighted in the elementary school curriculum, the addition and subtraction algorithms are fundamentally dependent on a thorough comprehension of the numerical system. This emphasis prompted CNESCO (Centre National d'Études des Systèmes Scolaires) to propose a recommendation titled 'Associating the learning of operational techniques with the understanding of numbers' following the 2015 consensus conference on initial learning in elementary school (CNESCO, 2015, p. 16). This guideline emphasizes that the instruction of written operation procedures (like carrying over in addition) should offer students the chance to enhance their numerical comprehension. It underscores the interconnection between the numerical system and computational algorithms. In the context of this study, we aim to investigate this connection by leveraging historical materials.

The value of a historical perspective on the evolution of number systems is commonly recognized (Bartolini Bussi & Sun, 2018), and there is now some consensus on the potential of history in mathematics education (Clark et al., 2018; Fauvel & Van Maanen, 2000). Among the possible interactions between the historical approach and the classroom are possible roles as a replacement (for another activity), as a change of scenery for the students, and as cultural enlightenment as well as options such as history as a tool or as a learning objective in itself or the choice of direct historical inputs or of a session inspired only by history. Undoubtedly, an understanding of the progressions leading up to our modern numerical system (including simple tally systems, additive systems, multiplicative-additive systems, and positional decimal systems) sheds light on the distinctive features and the epistemological hurdles inherent in our current system, particularly from a student's perspective. Consequently, it is unsurprising that the incorporation of historical numeral systems is frequently recommended in mathematics education. This approach has an inherently epistemological objective because it transforms what is assumed to be familiar into something unfamiliar, effectively prompting students to reconsider their conventional understanding of mathematical knowledge, as noted by Clark et al. (2018). In France, the historical approach is currently integrated into the majority of sixth-grade textbooks, as evidenced by our recent analysis of 22 sixth-grade textbooks published between 2005 and 2021. Within these textbooks, excluding documentary pages or inserts, we identified no fewer than 33 exercises related to various ancient numeral systems. These exercises encompassed Egyptian hieroglyphic numeration (ten), Roman numeration (nine), Babylonian cuneiform numeration (seven), Chinese and Sino-Japanese numeration (five), Mayan numeration (one), and Greek numeration (one). These exercises were often accompanied by visual illustrations, effectively emphasizing the historical and cultural context of each numeral system. In this regard, textbooks help to recall the history of numbering systems from which our contemporary system has evolved (Ifrah, 2000). The system used in France, as well as in most countries around the world, known as the Indo-Arabic numeral system, bears witness to the history of civilizations around the Mediterranean Sea and the scientific exchanges among peoples. Other numbering systems have developed independently, such as in Latin America and Asia. Nevertheless, they all share the same epistemological foundation in that they were designed to address the need for quantifying quantities. Only 8 textbooks did not mention ancient numeration. Besides the potential cultural enrichment, the incorporation of numerous ancient numeral systems into school exercises is closely intertwined with mathematics education. These ancient systems of number representation can be linked, firstly, to learning objectives, as outlined by De Vittori (2022), which are specific to our numerical system and the challenges it poses for students (base 10, position). Secondly, they are associated with the role of history in mathematics instruction. The experimentation discussed in this paper focuses on an ancient Chinese numerical system, where numbers are represented using rods, a system that is relatively well-documented. This Chinese numbering system was chosen without any particular link to a cultural anchor among the students for whom it appears to be just as exotic as an ancient Egyptian or Mayan system. Only the didactic considerations and mathematics learning issues, are considered. This system is briefly outlined, as follows.

AN ANCIENT CHINESE NUMERAL SYSTEM

In China, the practice of writing numbers with rods is believed to have emerged around the 2nd century BC, as evident in the Suàn shù shū [book on calculations made with sticks] from this era (Anicotte, 2019). This numerical system also resurfaced during the Wang Mang period (9-23 A.D.), as noted by Eberhard-Bréard (2008), and persisted until at least the early 18th century. It follows a positional decimal system in which distinct place values are presented through alternating horizontal and vertical representations. As Anicotte (2019) highlights in his edition, the book on calculations originally featured numbers written in the vernacular language. For mathematical calculations, numerical values were depicted using sticks on a flat surface, with arithmetic operations conducted through the manipulation of these sticks, known as suànchóu. The representation of numerical digits was, as follows:

For units, hundreds, and all even powers of ten, sticks were aligned vertically, and a horizontal bar denoted digits greater than five. For tens, thousands, and all odd powers of ten, the sticks were aligned horizontally, with a vertical bar indicating digits exceeding five. For instance, 〡 ≣ ⊤ represents the number 146, and 〡⊤ represents the number 106. In ancient China, the numerical system did not incorporate zero, as observed in Anicotte's (2019) work. To avert confusion, a blank space was sometimes left in its place. Zero made its appearance much later in subsequent versions, appearing as a small circle, as evidenced in the arithmetic triangle (Eberhard-Bréard, 2008) published in 1303 by Zhu Shijie (1260-1320). Black rods are used for negative numbers instead of red ones. In writings dating from the 11th century onwards, these negative numbers were distinguished with a slash. This numerical representation system gained widespread use from the 13th century, particularly in the context of solving algebraic equations.

In the experiment outlined in this paper, we focused on the decimal nature of the Chinese rod number system and emphasized the distinct representation of even and odd digit ranks in the writing. We intentionally accentuated the ambiguity arising from the absence of an explicit zero. In the didactic introduction of the system to the students, we deliberately refrained from suggesting the introduction of spaces between digits when a number unit was absent. The activity sheet and exercises presented the digits in a simple contiguous manner, requiring students to independently determine whether to arrange the symbols vertically or horizontally. However, to facilitate this deduction, the chosen numerical values featured only a single missing rank. For a more comprehensive exploration of the Chinese rod number system and its associated learning challenges, readers can refer to Bartolini Bussi and Sun (2018).

PRESENT STUDY

The potential value of incorporating history into mathematics education is world widely acknowledged (Clark & al., 2016; Fauvel & van Maanen, 2000). It offers an intriguing approach by providing both cultural enrichment and an epistemological perspective. However, while extensive research has been conducted on the significance of how students perceive and engage with mathematics (Bråting & Pejlare, 2015; Butuner, 2015; Lim & Chapman, 2015), several questions remain regarding the impact on the learning process. In their international synthesis, Clark et al. (2018) specifically seek empirical evidence to determine if students benefit from the use of mathematics history within the classroom. This article, using the introduction of Chinese numeration in a sixth-grade context as an example, addresses the following research question:

RQ. Does history of mathematics help students who underperformed on mathematics tests to improve their skills?

THEORETICAL FRAMEWORK

Interactions between history and student contemporary learning are complex but a good example of a framework that tries to consider this complexity can be found in the hermeneutic approach developed by Jahnke (2014). This pedagogical approach can be succinctly characterized by six principles, as delineated by Jahnke and discussed in Fried et al. (2016):

1. Students delve into a historical source after gaining a solid grasp of the corresponding mathematical concept in its modern form and perspective.
2. Students research and explore information related to the context and biography of the author.
3. The unique historical aspects of the source are preserved to the greatest extent possible.
4. Students are encouraged to generate free associations and creative insights.
5. The teacher emphasizes the need for well-reasoned arguments, while also allowing for diverse interpretations that need not be universally shared.
6. The historical comprehension of a concept is juxtaposed with its contemporary interpretation, fostering a process of reflection.

Hence, the hermeneutic approach can be perceived as a progression culminating in the sixth and ultimate step: a juxtaposition between contemporary and historical interpretations of mathematical concepts. At various levels, the activities found in French textbooks align quite effectively with the approach advocated by Jahnke. Typically, school exercises are introduced after the lesson, serving as fresh scenarios intended to reinforce the understanding of concepts. The activity presented in this article is consistent with these principles as well. It entails engaging with authentic historical resources to aid in contextualizing previously acquired mathematical knowledge. One possible research could be to assess the extent to which the historical approach contributes to the reactivation of mathematical concepts in an unconventional and disorienting context, as suggested by Barbin et al. (2020). This brings us to the issue of skills assessment and testing in schools.

From a methodological point of view, van den Linden (1986) reminds us of that item response theory (IRT) emerged in the 1950s and 1960s as a response to classical test theory. In contrast to classical test theory, which concentrates on random representative samples, IRT centers its attention on individuals' responses to individual test items. These responses are treated as the result of a stochastic experiment in which the likelihood of a particular response is contingent upon a set of parameters. These parameters can be related to the person or the item, such as its difficulty. The model proposed by Rasch (1960) indeed comprises only one parameter, which is difficulty. It posits that items have varying levels of difficulty but always the same discriminative power.

According to Rasch, the theoretical distribution of probabilities of correct responses to each item follows a logistic function. Since all items share the same distribution shape, the difficulty of a specific item is solely represented by the leftward (easier) or rightward (more difficult) shift of the abscissa of the inflection point of the curve. The modeling process involves determining these different difficulty levels for all items. Once the questions are modeled, one can also determine each student's proficiency level on the test, using Rasch unit known as a "logit". When a student's ability matches the item's difficulty, the probability of success is equal to 0.50. Rasch's model is widely used in the field of education, particularly for large-scale assessments like PISA (see OECD, 2009, chapter Rasch model), but also in medium and even small-scale empirical research. This enables a measure of students' skills that is relatively independent of contexts and local variations. In all the analyses of our experiment, we will also remain at the level of the statistical model. The main benefit of this approach is to derive conclusions that are as free as possible from a specific implementation or scale. The degrees of difficulty for various tasks and educational objectives are, therefore, guaranteed by the modeling itself, rather than depending on particular empirical values or subjective assessments. One can reasonably anticipate that the suggested activity, and possibly other activities following a similar model, will yield similar outcomes for students, irrespective of the context.

PARTICIPANTS

The experiment took place in France across four separate sixth-grade classes, each taught by a different teacher. In total, 108 students (comprising 63 girls [58%] and 45 boys [42%]) participated in either the entirety or portions of the experiment. Due to the COVID-19 pandemic, a few students were occasionally absent during certain phases of the experiment. However, a total of 86 students successfully completed all exercises and activities (n=86).

METHOD

Classroom Material

The experiment was structured into three sections (see **Appendix**). The initial segment encompassed standard mathematical exercises, aiming to assess how the integration of the history of mathematics correlated with the usual learning objectives (De Vittori, 2022). Only the items from questions 1, 2, and 3 are kept in the study because of their consistency with these specific learning contents. The second segment involved the history-centered activity, encompassing an introduction to the Chinese number system along with illustrative examples. The concluding part consisted of an assessment of students' comprehension of the Chinese number system as presented during the activity. The initial exercise sheet was assessed during a session before the activity, while the activity itself and the second worksheet were both completed within a single one-hour session. Data are collected on both the initial exercise sheet and the second one, thereby forming a pre- and post-test for the activity. The implementation of the activity is left up to the teacher. Indeed, the objective here is to assess the impact of an activity of the same kind as those found in school textbooks; that is, an all-in-one tool designed to be utilized without the need for additional input. For the purpose of IRT analysis, all tasks and responses were binarized using a stringent criterion, assigning a value of 1 for complete success and 0 otherwise. Items within the first exercise sheet featuring standard mathematical tasks were designated as M1a, M1b, M2a, M2b, and so forth. Similarly, items involving historical aspects in the final exercise sheet were labeled as H1a, H1b, H2a, H2b, and so on. The activity involved a dialogue with the teacher, while the two worksheets were individually completed by each student.

Worksheet 1: Regular School Mathematics/Pre-Test

The initial worksheet consists of three questions, totaling 12 items. The first question pertains to number dictation and serves to assess students' proficiency in the contemporary number system. In line with the learning challenges outlined in the preceding section, the chosen numbers are intended to spotlight potential weakness. As such, six items incorporate zeros in specific ranks (2,305, 10,100, 30,095) and large numbers (215,230, 6,800,000, 45,900,030). Subsequently, question 2 and question 3 relate to calculation algorithms. In question 2, students are tasked with adding three decimal numbers. Aiming to assess their comprehension of the various place values in our decimal system, the selected numbers increase in complexity. The initial two decimal numbers (3.29+1.05) have decimal portions of equal length. Subsequently, the decimal parts have varying lengths (66.7+2.42). Ultimately, the addition of an integer and a decimal (786+8.6). Question 3 follows a similar structure but focuses on subtraction (66.4-21.3), (24.1-0.25), and (2,043-22.2).

History-Based Activity

The discovery activity comprises two pages. The first page provides an explanation of the historical context and the mechanics of the Chinese number system. It presents both sets of digits for even and odd place values. At the bottom of the page, two examples, in addition to any that the teacher may suggest during the classroom session, are provided. These examples, namely 167 and 107, highlight the challenge arising from the absence of zero and the alternating symbols between digit ranks. The second page, in alignment with the format commonly found in educational textbooks that explore ancient numeration systems, involves converting from modern decimal notation to the ancient Chinese system, progressively dealing with larger numbers (12, 33, 46, 332, 467, and 5,678). Subsequently, a reverse encoding task is presented for the numbers 51, 81, and 457. This transition from the Chinese system back to our numbering system offers an opportunity to revisit the concept of zero, particularly in the context of the two numbers, 40 and 409 for which it is explicitly clarified that they are both less than 1000 to eliminate any potential ambiguity. The final part of the activity involves the reading of an original manuscript dating back to the 7th century, written in the Chinese numerical system. It features the numbers 9, 18, 27, 36, 45, 54, 63, 72, and 81. Students are tasked with identifying these numbers and recognizing the outcomes of multiplying by 9. The entire discovery activity takes place through classroom dialogue, allowing students to engage in discussions with both their peers and the teacher.

Worksheet 2: Evaluation/Post-Test

Following the completion of the activity, an assessment is provided to the students. This assessment sheet is based on the tasks encountered during the exploration of the Chinese number system. Each task on the assessment sheet specifies the ranges to which the numbers belong, and students are allowed to refer to the activity sheet throughout the evaluation. The assessment consists of a series of five questions, with each question containing two to four items, resulting in a total of 13 items. In the first question, students have to write the numbers 73, 221, and 6,789 in Chinese, while the second question asks them to rewrite the numbers 42, 50, 346, and 306 in our numerical system. These initial exercises involve applying the knowledge acquired during the activity and progressively include larger numbers that intentionally incorporate zeros. Questions 3, 4, and 5 present different challenges related to the concepts of the successor and predecessor of an integer. Connected to the ideas of addition and subtraction, the aim is to investigate whether the utilization of an alternative decimal system also facilitates comprehension of the fundamental rule of one-for-ten exchange. In question 3, students are tasked with writing, using the Chinese system, the successor of 23 and then the successor of 29. Similar scenario is presented in exercise 4 and exercise 5, which feature predecessors of 23 and 80, and of 6,789 and 6,780, respectively.

Examples of Students' Answers

As our study is intended to prepare for large-scale experimentation, it mainly focuses on quantitative approaches. Nevertheless, we present below a few examples of student productions in which some common difficulties can be observed. The main purpose of the activity is to work on the positional aspect of our numbering system, and one can see this learning issue clearly activated in some of the students' productions.

The alternating of different ranks and the absence of zero in the Chinese numbering system is quite disturbing at first for students. Some will therefore try to write numbers in a comparison with what they know about numbers in the French system. For example, a student (**Figure 1**) correctly re-encodes the numbers and identify the hundreds and ones ranks in Chinese, but he still writes 49 instead of 409.

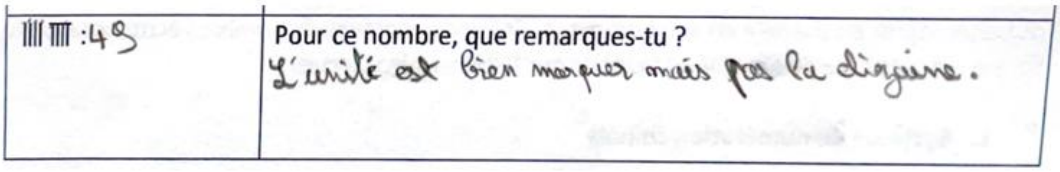

Figure 1. Student trying to deal with absence of zero (answer to asked question "What do you notice?": "Ones are well written but not tens") (Photo by G. Louaked)

For the same question, another student does not answer, considering that the Chinese number is simply not correct (**Figure 2**).

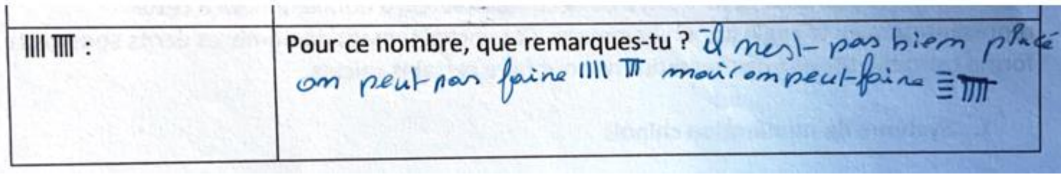

Figure 2. Student trying to deal with absence of zero (answer to asked question "What do you notice?": "It is not well placed. One cannot do IIII ╥ but one can do ≡╥ instead") (Photo by G. Louaked)

Other students reintroduce the zero, but we sometimes see confusion in the order of the different ranks. A student writes 04 instead of 40 even though he has understood the Chinese system (**Figure 3**).

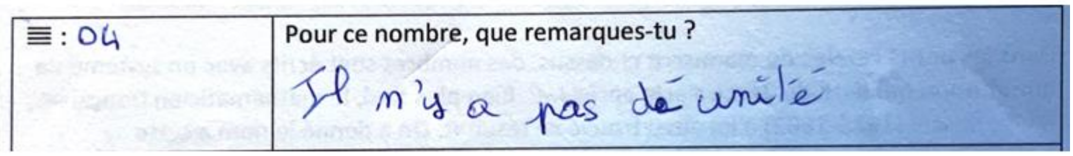

Figure 3. Student swapping tens & units in case of a zero (answer to asked question "What do you notice?": "There is no unit") (Photo by G. Louaked)

Another hesitates between 409 and 904 for the question that was mentioned previously (**Figure 4**).

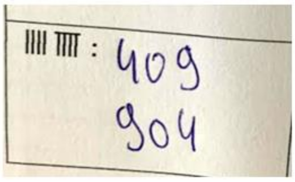

Figure 4. Student understanding well zero but hesitant about order (Photo by G. Louaked)

Finally, goes so far as to introduce a zero into Chinese system (**Figure 5**), a bit like what appears historically.

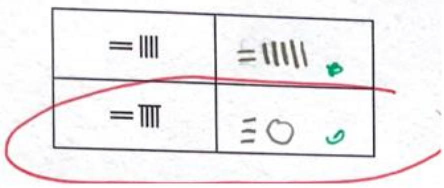

Figure 5. Student creating a zero in Chinese system, writing ≡0 instead of ≡ solely (Photo by G. Louaked)

As we can see from these examples, which are merely illustrative of a few extreme cases, the introduction of the Chinese numbering system destabilizes students' knowledge, forcing them to mobilize it in another way. There is no new knowledge about numbers here. The experiment therefore fits well into the hermeneutic approach recalled previously. With the Chinese number system, the goal is to get students to re-examine what might appear familiar.

ANALYSES

The data were processed in the *R* environment (R Core Team, 2022) with the *psych* package (Revelle, 2021) for data consistency checking, with the *ltm* package (Rizopoulos, 2006) for Rasch modeling, and with the *effsize* package (Torchiano, 2020) to compute the effect size index.

Results

Out of the 25 items, the overall average score is 19.16, with a standard deviation of 4.95. This outcome aligns with the aim of developing a classroom activity designed to be seamlessly integrated into regular instruction rather than serving as a certification-focused test (**Table 1** and **Table 2**). The results of student's t-test between girls and boys are not significant (t=0.061832; df=78.767; p-value=0.9509). Items were filtered based on point biserial correlation with total score item excluded (**Table 1** and **Table 2**). The literature (Slepkov et al., 2021) suggests a cutoff value of 0.2. Therefore, items M1a and H2a were removed. For the 23 remaining items, the consistency was good, and the presence of sufficient interitem correlations was ensured by a Cronbach's alpha of 0.89. The parallel analysis indicates one principal component and one factor, which confirms the uni-dimensionality of the data. Modeling with a Rasch model was therefore possible.

Table 1. Mean (*m*), standard deviation (σ), & point biserial correlation with total score item included (*PBi inc*)

Item	M1a	M1b	M1c	M1d	M1e	M1f	M2a	M2b	M2c	M3a	M3b	M3c
m	.97	.97	.91	.94	.78	.80	.91	.87	.70	.90	.66	.59
σ	.18	.18	.29	.24	.42	.40	.29	.34	.46	.31	.48	.49
PBi inc	.20	.43	.39	.47	.58	.69	.49	.55	.63	.29	.43	.51

Table 2. Mean (*m*), standard deviation (*σ*), & point biserial correlation with total score item excluded (*PBi ex*)

Item	H1a	H1b	H1c	H2a	H2b	H2c	H2d	H3a	H3b	H4a	H4b	H5a	H5b
m	.92	.92	.86	.90	.71	.77	.49	.78	.72	.83	.40	.58	.30
σ	.28	.28	.35	.31	.45	.42	.50	.42	.45	.38	.49	.50	.46
PBi ex	.40	.49	.37	.13	.32	.38	.47	.51	.65	.62	.40	.55	.47

Rasch Modeling & Item Selection

The fit of Rasch model was checked by bootstrap using the dedicated function in the *ltm* package (Rizopoulos, 2006). A p-value of 0.56 (greater than 0.05) for this model indicates a good overall fit. Items M2c and H5b are at the edge of the model, so we will avoid giving them a too significant weight in the analyses that follow.

To allow for a comparison of students' abilities between the mathematics and history parts, the items were paired by difficulty level following the method suggested by Wright (1993). Data from both tests are combined and analyzed together. Item calibrations are thus completely equated because they are all expressed at once on one common linear scale. One can then select pairs of items showing the same difficulty level. To constitute a pair, the difference in difficulty level between an M item and an H item must not exceed 0.2 in order to remain above the absolute mean error (Nurul Hafizah et al., 2020). Due to the limitations pointed out above, items M2c and H5b were not considered. By applying this process, we obtained six pairs of items of equivalent difficulty: M1c-H1a, M2a-H1b, M2b-H1c, M1e-H2c, M1f-H3a, and M3c-H5a (see **Table 3** for an overview of all items). Score means of participants on these sub-tests are almost the same 4.86 for M items, respectively 4.82 for H items, over 6. There is therefore no overall added value for all students (Wilcoxon-Mann-Whitney test p-value 0.5476 not significant), but there could be some, differentiated according to each student's proficiency in mathematics.

Table 3. Item difficulty (*diff*) in Rasch model

Item	M1b	M1c	M1d	M1e	M1f	M2a	M2b	M2c	M3a	M3b	M3c	
diff	-4.13	-2.90	-3.51	-1.62	-1.80	-2.90	-2.46	-1.06	-2.74	-0.85	-0.45	
Item	H1a	H1b	H1c	H2b	H2c	H2d	H3a	H3b	H4a	H4b	H5a	H5b
diff	-3.08	-3.08	-2.33	-1.21	-1.53	0.10	-1.62	-1.21	-2.00	0.59	-0.39	1.11

Added Value of History-Based Activity

To study a potential differentiated added value of the history of mathematics, the M items and the H items of each pair were separated and new Rasch modeling on each of these subtests was performed (**Figure 6**).

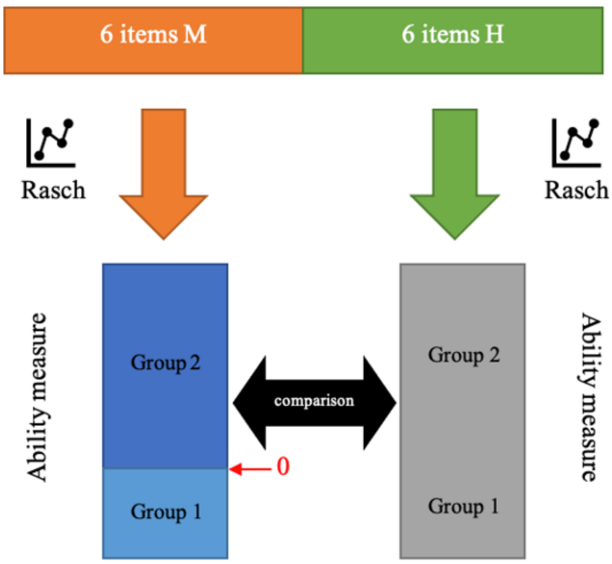

Figure 6. Protocol for measuring differentiated added value of historical approach (Figure by T. De Vittori)

To determine whether there was a differentiated benefit according to students' ability levels, two subsets of students were analyzed: the first one with students whose abilities, in Rasch modeling, were less than zero on the mathematics M items (23 students) and the second one with those whose ability was greater than zero (63 students). This measure of each student's ability within each group was then compared to that obtained on the history-based H items in the second worksheet.

For the first group, a significant increase (Wilcoxon-Mann-Whitney test p-value 0.003) in observed abilities of +0.67 logit in mean was measured. The corresponding Cliff delta index yielded a 0.54 large effect size. For the second group, a nonsignificant (Wilcoxon-Mann-Whitney test p-value 0.368) decrease in observed abilities of -0.21 logit in mean was measured. It appears that there is an effect observed among students who are the least proficient in mathematics.

As the study involves subsets, it is necessary to check whether the observed effect is not attributed, either wholly or partially, to the regression to the mean (RTM). RTM effect, first discussed by Sir Francis Galton in 1877, is a statistical phenomenon that occurs when studying certain subsets of data (Barnett et al., 2004). It is characterized by the tendency for extreme or outlier values within a subset to move closer to the overall average when measured again. This effect occurs because extreme values are often the result of random fluctuations or measurement error. On subsequent measurements, they are more likely to fall closer to the mean. It is important to be aware of this phenomenon when analyzing subsets, as it can lead to misleading conclusions if not properly accounted for in statistical analysis. As data in educational assessment is often discrete, ordinal, and poorly approximated by a normal distribution, Furrow (2019) suggests calculating numerical simulations of a null model to find the predictable impact of RTM. In order to create a null hypothesis (H0: "Measured effect is only due to RTM"), one can use permutations of the original data by randomly swapping pre- and post-test results (see more theoretical references in Furrow, 2019). This will preserve key relationships in the data and RTM effect will remain across all simulated datasets. Relevant statistics is calculated for each permutation. According to all principles, we create 10,000 permutations of our dataset, select less proficient students, and calculate mean change between pre- and post-test (**Figure 7**).

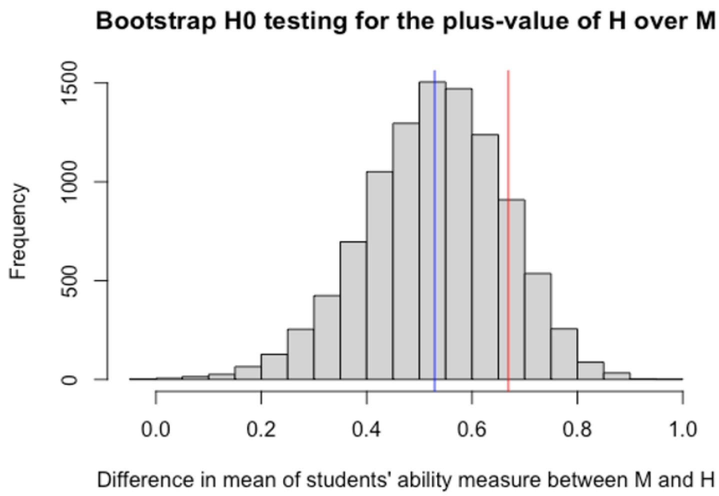

Figure 7. RTM effect estimation by bootstrap on 10,000 H0-type sets (RTM effect mean is +0.53 logit [in blue] & experimental value is +0.67 logit [in red] &least proficient students subset only) (Graph by T. De Vittori)

The overall RTM estimation by the permutation method is +0.53 logit within a 95% confidence interval [0.25; 0.77]. The experimental value +0.67 logit is not outside the confidence interval, so H0 cannot be rejected, which means that the observed effect cannot be attributed to the added value of the history-based approach. Corrected Cliff delta effect size is 0.10 (negligible).

Discussion

The experiment just presented was constructed to encourage students to think about numbering systems. For this purpose, among the many numbering systems that have been developed throughout history, an ancient Chinese numbering system was chosen.

This system shares enough similarities with our contemporary system to fit into students' school curricula. However, it also presents several differences, such as its unique writing system, which can facilitate an epistemological shift. The utilization of a historical source compels students to reevaluate their mathematical knowledge, encouraging them to manipulate both their own conceptions and those derived from the past (De Vittori, 2018). This approach is intended to promote learning, as suggested by Jahnke (2014). Both in the ordinary mathematics part and in the part with historical material, the learning issues are the same. By selecting items of equivalent difficulty in each of the two parts, we were able to assess whether the historical approach creates a different situation in which students better express their mathematical skills or not. The analysis of the data does not show any overall benefit for the students. This result can be compared with Butuner's (2015) meta-analysis, which reports at best only a weak effect of the use of history of mathematics on students, particularly concerning mathematics learning. For students who are generally the least proficient in mathematics, our study tends to show an added value of history, which can be largely explained by RTM. However, this result should be put into perspective by remembering that the scores obtained by the students during this experiment were very high. As we stated in the beginning of the data analysis, students were generally successful in these activities and exercises. The overall success rate is 0.77 (19.16 over 25, combining all M and H items), and considering only the paired items, the success rate is around 0.80 (4.8 over 6). The very principle of RTM is a tendency for the most extreme results to move closer to the mean. This phenomenon is certainly unfortunate if the average is low, but it may be desirable if the average is high. A simple way of anticipating the magnitude of RTM is to study the correlation coefficient ρ between pre- and post-test. Indeed, Kelly and Pride (2005) remind us that RTM effect is proportional to $(1-\rho)$. The poor correlation between our two tests ($\rho=0.43$, Spearman correlation method, paired items only) suggests a strong reshuffling of students' results. We can thus hypothesize that the history of mathematics creates a situation that is sufficiently new to, in some way, reset students' skills. In this new situation, the students have no experience and therefore no presuppositions (good or bad) about their skills. RTM then benefits students by reducing inequalities and enabling everyone to succeed, especially the least proficient. In this respect, our study is similar to that of Lim and Chapman (2015), who showed a positive effect of mathematics history on students' self-confidence. The authors also pointed out that this effect occurred mostly in the short term and tended to disappear after a few months. This could be explained by the observation that, when returning to regular school practices, the correlations between students' results tend to strengthen again. Our experiment was solely focused on mathematical learning content and did not evaluate the psycho-emotional aspects, which would necessitate a separate and dedicated study. We can therefore go no further in analyzing the medium- or long-term effects of historical approach. This leads us to present some limitations of our study.

Limitations of the Present Study

Initially, although our experiment's design aligns with typical textbook content, it remains unique and may potentially benefit from a novelty effect, unrelated to the historical material. We have not quantified this potential bias. To confirm the validity of our findings, it is advisable to replicate our results using similar activities across different classes following the same protocol. Furthermore, a longitudinal examination over time involving multiple activities with the same group of students could provide valuable insights.

The second limitation of our experiment is about to the representativeness of our sample. Our experiment was conducted within a single secondary school with different teachers for each class. Although the school is generally regarded as representative, potential biases may arise from the tested students' profiles. We have no real reason to believe that this had an impact on our results, but complementary research by randomized controlled trials (RCTs) would be desirable.

CONCLUSIONS

In this study, we tried to highlight indicators of a potential effect on mathematics performance of a history-based activity. The objective was to establish, within the framework of a pilot study, the first predicable values in statistical analyzes on large data sets. We created and implemented an activity based on an ancient Chinese number system, and we assessed students' skills on certain mathematical tasks before and after the activity. The abilities were evaluated *via* Rasch modeling and then compared.

This analysis does not highlight any overall improvement, but it shows a very large effect of RTM. Considering a context, where the averages are high and thus denote good success, it might change the usual interpretation. Perhaps this reshuffle is precisely, where relevance of historical approach for mathematical learning lies. A hypothesis for further research is then that the introduction of a historical perspective induces a decorrelation of student results between their initial level in mathematics and that, which they show during this type of activity. This reset can be a positive point and a way to understand and statistically measure the expected epistemological disorientation ("dépaysement épistémologique" in French) effect claimed by part of the international community involved in the use of history in mathematics education (Guillemette, 2018). In our dataset (paired items only), the per quartile analysis of RTM effect (R script available online, De Vittori, 2023) is ranging from +0.52 logit for the first quartile to -0.35 logit for the fourth (**Figure 8** and **Table 4**).

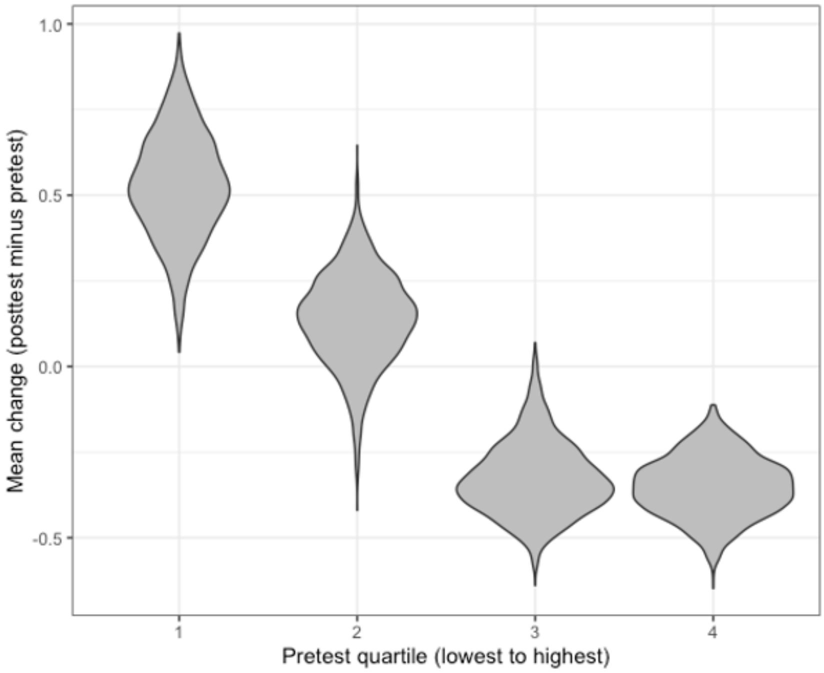

Figure 8. Boxplot of RTM effect for each quartile, in logit (Graph by T. De Vittori)

Table 4. RTM effect mean in logit & 95% confidence interval lower & upper bound for each quartile

Quartile	RTM effect mean	95% confidence interval lower	95% confidence interval upper
Q1	.52	.21	.82
Q2	.13	-.20	.39
Q3	-.33	-.52	-.10
Q4	-.35	-.53	-.19

According to results, RTM effect is non-negligible and might prove a true epistemological disorientation. It would obviously be preferable to have a positive effect on all students, but this will require other studies on the development of skills in the medium and long term. Positive or not, it will be necessary in any case to control RTM effect, in particular by using an RCT approach as suggested in the literature (Barnett et al., 2004). This will be one of the main guidelines of our next large-scale studies.

Author contributions: TDV: designed experiment & carried out statistical analyses & interpretations & **GL & M-PV:** implemented experiment. All authors approved the final version of the article.

Funding: This article was supported by the Institut National Supérieur du Professorat et de l'Éducation de l'académie de Lille–Hauts-de-France, France.

Acknowledgements: The authors would like to thank editor & anonymous reviewers for their valuable comments.

Ethics declaration: The authors declared that analysis was conducted fully anonymized & there exists no possibility to reconstruct identities of participants from this study. The authors stated that ethical review & approval was not required study on participants in accordance with the General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament).

Declaration of interest: The authors declared no competing interest.

Data availability: Raw data available on Zenodo permanent deposit: https://doi.org/10.5281/zenodo.6392984

REFERENCES

Anicotte, R. (2019). *Le livre sur les calculs effectués avec des bâtonnets: Un manuscrit du -IIe siècle excavé à Zhangjiashan* [*The book of calculations made with sticks: A 2nd-century manuscript excavated in Zhangjiashan*]. Presses de l'Inalco. https://doi.org/10.4000/books.pressesinalco.18815

Barbin, E., Guillemette, D., & Tzanakis, C. (2020). History of mathematics and education. In S. Lerman (Ed.), *Encyclopedia of mathematics education*. Springer. https://doi.org/10.1007/978-94-007-4978-8_69

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215-220. https://doi.org/10.1093/ije/dyh299

Barr, D. C. (1978). A comparison of three methods of introducing two-digit numeration. *Journal for Research in Mathematics Education, 9*(1), 33-43. https://doi.org/10.2307/748958

Bartolini Bussi, M. G., & Sun, X. H. (2018). *Building the foundation: Whole numbers in the primary grades. The 23rd ICMI study*. Springer. https://doi.org/10.1007/978-3-319-63555-2

Baturo, A. (2000). Construction of a numeration model: A theoretical analysis. In J. Bana, & A. Chapman (Eds.), In *Proceedings of the 23rd Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 95-103). https://merga.net.au/Public/Publications/Annual_Conference_Proceedings/2000_MERGA_CP.aspx

Bednarz, N., & Janvier, B. (1982). The understanding of numeration in primary school. *Educational Studies in Mathematics, 13*(1), 33-57. https://doi.org/10.1007/BF00305497

Bråting, K., & Pejlare, J. (2015). On the relations between historical epistemology and students' conceptual developments in mathematics. *Educational Studies in Mathematics, 89*(2), 251-265. https://doi.org/10.1007/s10649-015-9600-8

Butuner, S. O. (2015). Impact of using history of mathematics on students mathematics attitude: A meta-analysis study. *European Journal of Science and Mathematics Education, 3*(4), 337-349. https://doi.org/10.30935/scimath/9442

Chambris, C., & Tempier, F. (2017). Dealing with large numbers: What is important for students and teachers to know? In T. Dooley, & G. Gueudet (Eds.), *Proceedings of the 10th Congress of the European Society for Research in Mathematics Education* (pp. 245-252). DCU Institute of Education and ERME. https://hal.science/CERME10

Chesné, J.-F., & Fischer, J.-P. (2015). Les acquis des élèves dans le domaine des nombres et du calcul à l'école primaire. Rapport pour la conférence de consensus nombres et opérations: Premiers apprentissages à l'école primaire [Students' achievements in the area of numbers and calculation in primary school. Report for the consensus conference numbers and operations: First learning in primary school]. *CNESCO*. https://www.cnesco.fr/wp-content/uploads/2015/11/Acquis-des-élèves.pdf

Clark, C., Kjeldsen, T. H., Schorcht, S., Tzanakis, C., & Wang, X. (2016). History of mathematics in mathematics education: Recent developments. In L. Radford, F. Furinghetti, & T. Hausberger (Eds.), *Proceedings of the 2016 ICME Satellite Meeting–HPM 2016* (pp. 135-179). IREM de Montpelier.

Clark, K. M., Kjeldsen, T. H., Schorcht, S., & Tzanakis, C. (2018). *Mathematics, education and history: Towards a harmonious partnership*. Springer. https://doi.org/10.1007/978-3-319-73924-3

CNESCO. (2015). *Conférence de consensus. Nombres et opérations: Premiers apprentissages à l'école primaire. Recommandations du jury* [*Consensus conference. Numbers and operations: First learning in primary school. Jury recommendations*]. https://www.cnesco.fr/wp-content/uploads/2015/11/Recommandations-du-jury.pdf

De Vittori, T. (2018). Analyzing the use of history in mathematics education: Issues and challenges around Balacheff's cKȼ model. *Educational Studies in Mathematics, 99*(2), 125-136. https://doi.org/10.1007/s10649-018-9831-6

De Vittori, T. (2022). Relevance of a history-based activity for mathematics learnings. *Discover Education, 1*(1). https://doi.org/10.1007/s44217-022-00010-1

De Vittori, T. (2023). Regression to the mean (RTM) effect calculation by random data permutations. *Zenodo*. https://doi.org/10.5281/zenodo.8344700

Eberhard-Bréard, A. (2008). Mathematics in China. In H. Selin (Ed.), *Encyclopedia of the history of science, technology, and medicine in non-Western cultures*. Springer. https://doi.org/10.1007/978-1-4020-4425-0

Fauvel, J., & Van Maanen, J. (2000). *History in mathematics education*. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47220-1

Fried, M. N., Guillemette, D., & Jahnke, H. N. (2016). Theoretical and/or conceptual frameworks for integrating history in mathematics education. In L. Radford, F. Furinghetti, & T. Hausberger (Eds.), *Proceedings of the 2016 ICME Satellite Meeting of the International Study Group on the Relations Between the History and Pedagogy of Mathematics* (pp. 211-230). IREM de Montpellier. https://hal.science/HPM2016/public/hpm2016_eproceedings_final.pdf

Furrow, R. E. (2019). Regression to the mean in pre-post testing: Using simulations and permutations to develop null expectations. *CBE–Life Sciences Education, 18*(2), le2. https://doi.org/10.1187/cbe.19-02-0034

Guillemette, D. (2018). History of mathematics and teachers' education: On otherness and empathy. In K. Clark, T. Kjeldsen, S. Schorcht, & C. Tzanakis (Eds.), *Mathematics, education and history. ICME-13 monographs*. Springer. https://doi.org/10.1007/978-3-319-73924-3

Houdement, C., & Tempier, F. (2019). Understanding place value with numeration units. *ZDM Mathematics Education, 51*(1), 25-37. https://doi.org/10.1007/s11858-018-0985-6

Ifrah, G. (2000). *The universal history of numbers*. Wiley.

Jahnke, H. N. (2014). History in mathematics education. A hermeneutic approach. In M. Fried, & T. Dreyfus (Eds.), *Mathematics & mathematics education: Searching for common ground. Advances in mathematics education.* Springer. https://doi.org/10.1007/978-94-007-7473-5_6

Kelly, C., & Price, T. D. (2005). Correcting for regression to the mean in behavior and ecology. *The American Naturalist, 166*, 700-707. https://doi.org/10.1086/497402

Lambert, K., & Moeller, K. (2019). Place-value computation in children with mathematics difficulties. *Journal of Experimental Child Psychology, 178*, 214-225. https://doi.org/10.1016/j.jecp.2018.09.008

Lim, S. Y., & Chapman, E. (2015). Effects of using history as a tool to teach mathematics on students' attitudes, anxiety, motivation and achievement in grade 11 classrooms. *Educational Studies in Mathematics, 90*(2), 189-212. https://doi.org/10.1007/s10649-015-9620-4

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Lawrence Erlbaum. https://doi.org/10.4324/9781410602589

Nurul Hafizah, A., Zamalia, M., & Adzhar, R. (2020). Rasch rating scale item estimates using maximum likelihood approach: Effects of sample size on the accuracy and bias of the estimates. *International Journal of Advanced Science and Technology, 29*(4s), 2526-2531.

OECD. (2009). *PISA data analysis manual: SAS*. OECD Publishing. https://doi.org/10.1787/9789264056251-en

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danemarks Paedogogiske Institut.

Revelle, W. (2021). *psych: Procedures for personality and psychological research*. https://cran.r-project.org/package=psych

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1-25. https://doi.org/10.18637/jss.v017.i05

Slepkov, A. D., Van Bussel, M. L., Fitze, K. M., & Burr, W. S. (2021). A baseline for multiple-choice testing in the university classroom. *SAGE Open, 11*(2). https://doi.org/10.1177/21582440211016838

Thanheiser, E. (2012). Understanding multidigit whole numbers: The role of knowledge components, connections, and context in understanding regrouping 3+- digit numbers. *The Journal of Mathematical Behavior, 31*(2), 220-234. https://doi.org/10.1016/j.jmathb.2011.12.007

Thomas, N. (2004). The development of structure in the number system. In M. J. Hoines, & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (pp. 305-312). Bergen University College Press. https://files.eric.ed.gov/fulltext/ED489178.pdf

Torchiano, M. (2020). effsize: Efficient effect size computation. *Zenodo*. https://doi.org/10.5281/zenodo.1480624

van der Linden, W. J. (1986). The changing conception of measurement in education and psychology. *Applied Psychological Measurement, 10*(4), 325-332. https://doi.org/10.1177/014662168601000401

Wright, B. D. (1993). Equitable test equating. *Rasch Measurement Transactions, 7*(2), 298-299.

APPENDIX

Note. Where useful, expected answers are given in green.

Worksheet (1)

Question 1

Table A1. Numbers dictation

2,305	10,100	30,095	215,230	6,800,000	45,900,030

Question 2

Table A2. Carries out following additions using column method

3.29+1.05	66.7+2.42	786+8.6

Question 3

Table A3. Carries out following subtractions using column method

65.4-21.3	24.1-0.25	2,043-22.2

Main Activity Sheet: A Chinese Numbering System from the 2nd Century A.D.

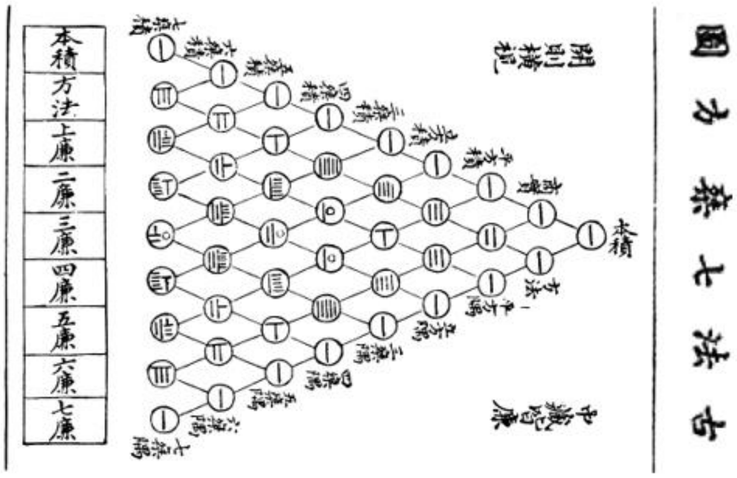

Figure A1. Pascal's triangle published in 1303 by Zhu Shijie (1260-1320) (Image: Wikipedia)

In the small circles of the manuscript in **Figure A1**, numbers are written using a numbering system that dates back to the 2nd century A.D. Much later, the French mathematician Blaise Pascal (1623-1662) also found this result. The triangle representation was named in his honor. This sequence of numbers written in this form is very useful to mathematicians for certain calculations.

Chinese numbering system

In the 2nd century A.D., the Chinese wrote numbers using a system that works, as follows: The **system is decimal**.

Digits of **units** and **hundreds** are represented by arranging rods, as shown in **Table A4**:

Table A4.

One	Two	Three	Four	Five	Six	Seven	Eight	Nine
𝍠	𝍡	𝍢	𝍣	𝍤	𝍥	𝍦	𝍧	𝍨

To make reading easier, the rods representing the **tens** and **thousands** digits are arranged differently (**Table A5**): For instance, 𝍠𝍮𝍦 represents the number 167 and 𝍠𝍦 represents the number 107.

Table A5.

One	Two	Three	Four	Five	Six	Seven	Eight	Nine
𝍩	𝍪	𝍫	𝍬	𝍭	𝍮	𝍯	𝍰	𝍱

First examples

Using the rules above, write the following numbers in 2nd century Chinese:

12: 𝍩𝍡

33: 𝍫𝍢

46: 𝍬𝍥

332: 𝍢𝍫𝍡

467: 𝍣𝍮𝍦

5,678: 𝍭𝍥𝍰𝍧

Reading an ancient Chinese manuscript

Write the following numbers using the French system, knowing that all the numbers are less than 1,000:

𝍭𝍠 : 51

𝍰𝍠 : 81

𝍣𝍭𝍦: 457

𝍬: 40

What do you notice about this number?

Expected answer: There is no zero for the units.

𝍣𝍨: 409

What do you notice about this number?

Expected answer: There's no zero for the tens.

The manuscript in **Figure A2** is a mathematical text from the 7th century. The last line gives a sequence of numbers. Write this list of numbers using the French system.

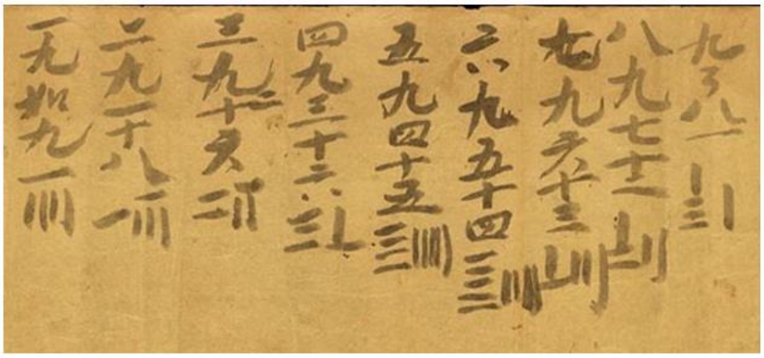

Figure A2. Manuscript (source: British Museum (Or.8210) International Dunhuang Project [idp.bnf.fr])

Table A6.

Numbers	9, 18, 27, 36 (error in manuscript, symbols flipped vertically), 45, 54, 63, 72, 81
Do you recognize sequence of numbers?	Expected answer: This is nine times multiplication table.

Worksheet (2)

Note. *For all this part, the table showing the correspondences between French and Chinese numbers is given to each student.*

Question 1

Table A7. Write following numbers using Chinese system

73: 𝍯𝍢	221: 𝍡 𝍪 𝍠	6,789: 𝍮𝍦𝍰𝍨

Question 2

Table A8. Write following numbers using French system, knowing that all numbers are less than 1,000

𝍬 𝍡 : 42	𝍭: 50	𝍢 𝍬 𝍥: 346	𝍢 𝍥: 306

Question 3

Table A9. Using Chinese system, write number that comes just after each of following numbers (all numbers between 10 & 99)

𝍪𝍣	𝍪𝍤
𝍪𝍨	𝍫

Question 4

Table A10. Using Chinese system, write number that comes just before each of following numbers (all numbers between 10 & 99)

𝍪 𝍡	𝍪 𝍢
𝍯𝍨	𝍰

Question 5

Table A11. Using Chinese system, write number that comes just before each of following numbers (all numbers between 1,000 & 9,999)

𝍮𝍦𝍰𝍧	𝍮𝍦𝍰𝍨
𝍮𝍦 𝍯𝍨	𝍮𝍦𝍰