Constructing a Norm for Children's Scientific Drawing: Distribution Features Based on Semantic Similarity of Large Language Models

Yi ZHANG[1]+, Fan WEI[2]+, Jingyi LI[3]+, Yan WANG[1], Yanyan YU[1], Jianli CHEN[1], Zipo CAI[1], Xinyu LIU[1], Wei WANG[1], Peng WANG[4], Zhong WANG[5]*

1. First Primary School, Fengtai, Beijing;
2. Fangzhuang Primary School, Fengtai, Beijing;
3. College of Elementary Education, Capital Normal University, Beijing;
4. Hepingli No.9 Primary School, Dongcheng, Beijing;
5. Beijing Doers Education Consulting Co., Ltd, Beijing

Note: '+'is the first author, '*' is the corresponding author

Abstract: The use of children's drawings to examining their conceptual understanding has been proven to be an effective method, but there are two major problems with previous research: 1. The content of the drawings heavily relies on the task, and the ecological validity of the conclusions is low; 2. The interpretation of drawings relies too much on the subjective feelings of the researchers. To address this issue, this study uses the Large Language Model (LLM) to identify 1420 children's scientific drawings (covering 9 scientific themes/concepts), and uses the word2vec algorithm to calculate their semantic similarity. The study explores whether there are consistent drawing representations for children on the same theme, and attempts to establish a norm for children's scientific drawings, providing a baseline reference for follow-up children's drawing research. The results show that the representation of most drawings has consistency, manifested as most semantic similarity>0.8. At the same time, it was found that the consistency of the representation is

independent of the accuracy (of LLM's recognition), indicating the existence of consistency bias. In the subsequent exploration of influencing factors, we used Kendall rank correlation coefficient to investigate the effects of "sample size", "abstract degree", and "focus points" on drawings, and used word frequency statistics to explore whether children represented abstract themes/concepts by reproducing what was taught in class. It was found that accuracy (of LLM's recognition) is the most sensitive indicator, and data such as sample size and semantic similarity are related to it; The consistency between classroom experiments and teaching purpose is also an important factor, many students focus more on the experiments themselves rather than what they explain. In addition, most children tend to use examples they have seen in class to represent more abstract themes/concepts, indicating that they may need concrete examples to understand abstract things.

Keywords: Children's drawings, Norm, LLM, Word2vec, Semantic similarity, Distribution

Corresponding author: Zhong WANG (0000-0003-2020-8450), Beijing Doers Education Consulting Co., Ltd, wangzhong@bdice.ac.cn

1. Introduction

Children's drawing is a commonly used way of examining children's cognitive and emotional status [1]. The reason for this is not only because drawing is a pleasurable activity for children [2], but also because its effectiveness in examining children's psychological activities is considered reliable and has been extensively studied in many fields. For example, from the perspective of individual learning, drawing is used to examine children's cognition of a certain scientific concept. Brechet et al. had over 250 French children draw their brains to understand how the "black box" of the brain works [3]. In terms of emotions

and interpersonal relationships, campus bullying is a topic of great concern to psychologists, so drawing has also been used in related research. For example, Marengo et al. analyzed more than 600 drawings of Italian children to understand this issue [4].

However, there are still some fundamental issues that have not been overcome in the current psychological analysis of children's drawings, which mainly focus on methodological perspectives. This is mainly reflected in two aspects: firstly, the issue of ecological validity. As drawing is a non-verbal representation, it must be digitized through effective encoding in order to be analyzed using statistical methods. Therefore, encoding method is currently the mainstream way to study children's drawing. For example, Tolsberg et al. used coding to understand the longitudinal development of math word problems among Estonian elementary school students [5]. The coding rule for that study is as follows: those without drawings are coded as 9, those depicted as pictorial are coded as 1, and those depicted as schematic are coded as 2. The main difference between 1 and 2 is whether the relationship between the things is revealed. Therefore, how to develop coding rules is a crucial issue. This issue is particularly prominent in learning science, as emotions and interpersonal relationships has a large number of scales or tools (such as bullying problems [6]), and some drawing studies on this issue can use these general tools [7]. However, many researchers in learning science area had to adopt self-developed coding rules. In a psychology journal, there were a total of 12 original studies on the topic of children's drawing research, of which only 4 studies did not involve self-developed coding at all (only 33.3%) [8].

The high proportion of self-developed coding naturally raises a question: how can these highly task dependent coding rules be transferred and generalized to other drawing studies? In other words, different studies have different tasks,

so how should other drawing studies refer to the results of these codes? Obviously, existing drawing research lacks a unified norm, which allows students to draw a certain field or object relatively freely under the condition of no tasks or fewer tasks, in order to fully expose their drawing preferences for that object in the naturally state. If this norm exists, different drawing tasks can be compared with it, then there is a common reference frame for drawing research on the same theme.

The second issue is more obvious: the understanding of drawing is one of the most prominent areas of individual heterogeneity, and different people's perspectives on the same drawing may be completely different [9]. Therefore, for studies do not utilize machines, consistency testing is an essential step - the same drawing must be identified by at least two researchers, and the results of these two identifications must pass consistency testing (such as Cohen Kappa coefficient [3] [5]). On the contrary, if machine learning models are used to identify drawings, this process be likely to omitted. Because obviously, the same model is based on the same algorithm, its generation results should be more consistent rather than researchers.

Based on the above two issues, this study aims to construct a norm of drawings for children's understanding of scientific themes, and to use the Large Language Model (LLM) to understand and identify children's drawings, attempting to provide a reference frame for future research on children's conceptual understanding of drawing. This study focuses on three questions: 1. Do children’s drawings have a consistent representation for the same drawing theme (such as whether the concept of solar eclipse is mostly expressed through light path diagrams)? 2. Is there a relationship between the consistent representations (if it exists) and the accuracy of drawing recognition of LLM? 3. What other factors are related to the consistent representations (if it exists)?

2. Method

2.1 Participants

Select a total of 1473 students from two primary schools in Beijing. Among them, there are 74 students in grade 4, and the rest are in grades 5 and 6. The male to female ratio is 1.01:1. The textbooks involved two versions [10] [11].

We excluded drawings that were unclear, incomplete, or clearly did not belong to the prescribed content, as well as drawings with inconsistent recognition results by LLM multiple times (see section 2.2.3 of this article), and finally included a total of 1420 drawings in the sample.

All children's personal information is strictly confidential, and the drawings were sealed after being photographed to ensure that personal information of students were not leaked. And informed consent from the participants' guardian was obtained. This research was approved by the Research Ethical Review Committee of Beijing Doers Educational Consulting Co., Ltd on March 13, 2024 (ECS-008).

2.2 Proceeding

2.2.1 Collection of drawings:

Firstly, all children draw in class, that is, they draw whatever they just learned (following the order of textbooks), without deliberately setting contents or tasks. Secondly, to ensure the consistency and interpretability of the data, all participating teachers in this study are required to strictly follow the textbook and strictly prohibit free design teaching scaffolds and guidance to prevent individual differences among teachers from interfering with the results. Thirdly, the prompt was: "We have just learned the knowledge of …, please use drawing to represent it". Other than that, no hints will be given to prevent any suggestion

or misleading to students. Fourth, ensure that all students can complete the drawing.

2.2.2 Drawings Preprocessing:

Firstly, number and take photos of all the drawings. Secondly, preprocessing the photos includes two parts: ①removing students' personal information such as names and student IDs; ②removing the theme/concept names (some students have the habit of writing the names of the themes/concepts on their drawings) to prevent prompting LLM. The solution is to use the "rubber stamp" function of Photoshop software (v.2018) to cover these names with surrounding images.

2.2.3 LLM identification of drawings:

Firstly, input the photos into LLM for identification. The prompt is "What is the specific content of this drawing? How did you see it from the picture?", and copy all the results and reasons generated by LLM into an Excel file for analysis.

In addition, to ensure the stability of LLM recognition, all drawings are independently recognized twice by LLM. Drawings with significantly different recognition results (such as the first recognition as a solar eclipse and the second recognition as a lunar phase) will undergo a third independent identification, and those with two occurrences of the three results will be considered as the result. If three different results occur, the drawing will be removed. The so-called 'independence' refers to deleting the current page after one recognition is completed and re recognizing it. Since the LLM used does not capture conversations with front-end users and place them in its pre training set, the above operation method can ensure the independence of the two recognitions.

This study used ChatGLM-4.0 to identify drawings. The reason for adopting this LLM is that its image recognition models mainly use convolutional neural networks (CNN) and recurrent neural networks (RNN) (derived from the self-introduction of this LLM), and these two models have been proven to be efficient and reliable in image recognition (because the two models are so famous and there are too numerous related studies, so this article will not list the literatures here). And there have been some studies published using this LLM for image recognition [12] [13], further proving its reliability in image recognition. And its language generation method adopts a model based on the transformer algorithm, which has high efficiency and reliability in language generation. More importantly, this LLM is open to individual users for free, making it more suitable for numerous primary school teachers to identify children's drawings, as mentioned above.

To verify that LLM's image recognition reasonings can indeed explain children's drawings representations, we counted the number of objects appearing in the drawings in all LLM image recognition reasoning (as long as one object appears in the drawings in these sentences, it is considered). The results showed that 77.4% (weighted average) of the reasoning statements for image recognition contained at least one object appearing in the drawings. This indicates that LLM is indeed describing the specific content of the drawings, rather than fabricating reasons. At the same time, in order to prevent LLM from generating image recognition reasons by reading text on the drawings, we randomly selected 16 drawings without any prompts (2 drawings for each theme/concept, and all drawings of the "solar system" concept have prompts, so this concept cannot be selected). The results showed that LLM described at least one object in all the 16 drawings.

2.3 Semantic Similarity Analysis

2.3.1 Algorithm

This study used the semi supervised Word2vec algorithm to generate word vectors, and then used the cosine similarity algorithm to calculate semantic similarity values.

The Word2vec is a deep learning algorithm for word embedding proposed by the Tomas Mikolov team at Google. Traditional spatial distance-based algorithms (such as Euclidean distance or Manhattan distance) in handling sentences with the same semantics but different representations usually have poor performance [14]. In contrast, the word2vec algorithm creatively proposes a negative sampling method, which calculates the conditional probability of the sampled sample coming from the corpus through context word frequency statistics, and then determines its polarity [15]. This method can better identify similar semantic sentences with different representations.

Cosine similarity is a widely used algorithm for calculating semantic similarity. Its main principle is to model the text as a term vector and calculate the cosine value between the term vectors to determine the degree of similarity between two texts [16]. The cosine similarity algorithm has significant advantages in semantic similarity calculation, with evidence showing that it has higher resolution than Jaccobi similarity [17].

By incorporating supervision into unsupervised algorithms, it has been widely applied in NLP semantic analysis and comparative research on cross class text learning between machines and humans, demonstrating its advantages [18].

2.3.2 Proceeding of Semantic Similarity Analysis:

Firstly, we collated all the statements generated by LLM for image recognition, match the statements with the image numbers one by one, and outputted the

results as an Excel file. Then, used the JIEBA tokenizer to segment the statements in this Excel file to words (removed all punctuation marks). Then we trained a semi supervised word2vec algorithm (dimension 100, context window size 5, negative sampling 5) to embed word vectors into the segmentation results, and then used the word vector embedding results to calculate the average word vector of each sentence. Finally, the cosine similarity algorithm was used to calculate the semantic similarity between all pairs of sentences, thereby forming a similarity matrix.

In addition, although the theoretical range of semantic similarity calculated by the cosine similarity algorithm should be [-1,1], all semantic similarity results in this article are greater than 0. Therefore, for the sake of simplicity in calculation, all calculations regarding the concentration of semantic similarity in the following text were defaulted the value range of [0,1] instead of [-1,1].

2.3.3 The Semi Supervised Algorithms

Due to word2vec being an unsupervised algorithm, supervised learning is clearly superior in most cases in NLP processing. Therefore, we introduce keywords to transform the above algorithm into a semi supervised learning method, thereby increasing the accuracy of word vector embedding.

The specific method was to first use the JIEBA tokenizer to segment the text to be analyzed and removed all punctuation marks. Secondly, used the TF-IDF algorithm to search for keywords in the text and assigned weights to these keywords. The number of keywords of each text was set at 10, because after repeated experiments, when the number exceeds 10 (such as 15), there will be a large number of high-frequency function words (such as "this", "yes", etc.) in the keywords. Thirdly, implemented semi supervision by adding Python statements. Finally, embed the generated word vectors into an Excel file to

calculate the average word vector for each sentence in the original text.

2.4 Distribution Analysis and Visualization

Used a heat map to represent the similarity values of the semantic similarity matrix for each theme/concept, and used a histogram to represent the specific distribution of similarity in [0,1].

Specifically, we divided all data in each similarity matrix into eight equal parts in order of size (similar to quartiles), and separately calculated the difference between the seventh point (87.5%, from the minimum) and the second point (12.5%), which represents the size of the region where 75% of the semantic similarity values in this matrix are concentrated within [0,1] (hereinafter referred to as the "75% Region"). The smaller the value, the higher the numerical concentration, which is used to examine whether the theme/concept has a consistent representation.

2.5 Exploration of Related Factors

Finally, we used Kendall rank correlation coefficient (Kendall - τ) to calculate the effects of three factors: Sample Size, Abstract Degree of Themes/Concepts, and Reasoning Degree of Themes/Concepts on the Accuracy, Semantic Similarity, and 75% Region of children's drawing representations.

There are two reasons for using this indicator (Kendall - τ): firstly, this article is only an observational study, not a control study, so exploring correlation is more suitable than establish the causal relationship [19]. Secondly, the related factor exploration in this article aims to understand the impact of a certain factor on 9 concepts, which means we only have a sample size of 9. The commonly used Pearson correlation coefficient is an indicator that strictly depends on sample size, but due to the limitations of researchers, we do not have the ability to

explore more concepts of children's drawings. The Kendall - τ is more suitable for dealing with low sample size correlation problems, so we chose this indicator to explore the relevant factors.

At the same time, we also examined the reproduction rate of children's experiments/phenomena taught in class (mainly for abstract themes/concepts) through the method of word frequency. Because we suspect that students may have difficulty grasping the connotations of abstract concepts, they may use experiments/phenomena taught in class to represent these concepts in drawings.

3. Results

3.1 Overall

The total number of samples is 1420, covering 9 scientific themes or concepts: Circuits, Solar eclipses, Boiling, Solar system, Increasing the carrying capacity of a boat, Life history of a plant, Buoyancy, Electromagnetics, Physical and chemical changes. The average recognition accuracy of LLM is 46.7%, and the median semantic similarity is 0.9885. In addition, the mean of the "75% Region" is 0.0035 (i.e. occupying 0.35% of [0,1]). The data is detailed in Table 1.

From the overall situation, it can be seen that LLM has a low recognition accuracy for image drawing content, but the semantic similarity concentration is high. This seems imply that there was a consistency preference in the representation of most of the drawings, but there is a possibility of "collective bias" in this preference (i.e. those children's drawing representations were inaccurate, but the way these erroneous representations were consistent).

Concepts	Sample size	Accuracy	Median value	75%
Circuit	267	0.843	0.9998	0.02
Solar eclipses	214	0.44	0.9999	0
Boiling	48	0	0.967	1.49
Solar system	201	0.65	0.9999	0

Increasing the carrying capacity of a boat	52	0.033	0.987	1.29
Life history of a plant	368	0.712	0.9999	0.01
Buoyancy	40	0.5	0.944	3.38
Electromagnetics	70	0.514	0.9995	0.12
Physical and chemical changes	160	0.513	0.9999	0

Table 1 Overall situation of the samples (where 75% refers to the 75% Region, with smaller values being more concentrated)

3.2 Results of every theme/concept

3.2.1 Circuit

The main learning content (experiment) of this concept is to light up a bulb to let students know that the circuit must be closed in order to work (i.e. form a closed loop).

The average accuracy of LLM image recognition is 84.3%, which is the highest among all nine themes/concepts, indicating that most students can grasp the correct representation method of this concept.

The median value of semantic similarity is 0.9998, and 75% Region is concentrated within 0.02% of [0,1], indicating that the vast majority of students have a high degree of consistency in the representation of the form of circuits. Considering the high accuracy, it indicates that the majority of students' drawings are correct and consistent. For specific details, please refer to Figures 2 and 3, where Figure 2 is a numerical analysis and Figure 3 is the “representative drawings”. The so-called "representative drawings" refer to a pair of drawings with semantic similarity located in the median of the 75% Region. As they are in the median, they should be the most representative of all drawings, the same applies below.

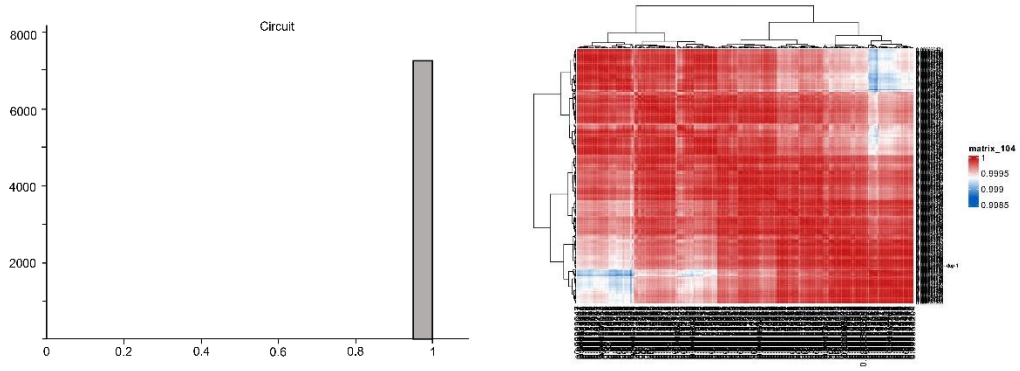


Figure 1: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Circuit". The heatmap on the right specifically refers to the semantic similarity data distribution of the ≥ 0.9 part (the same below)

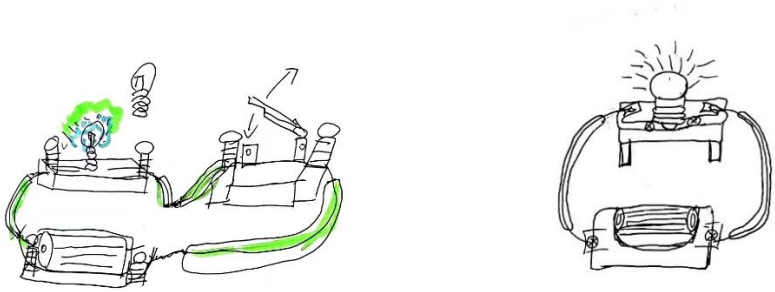


Figure 2: Representative drawing of the concept of "Circuit"

3.2.2 Solar eclipse

The main content of this concept is to understand the phenomenon and causes of solar eclipses, but it does not require representation using light paths.

The average accuracy of LLM image recognition is 44%, which is basically at the average level of nine themes/concepts.

The median value of semantic similarity is 0.9999, and 75% Region is concentrated within the 0.001% range of [0,1], indicating that the vast majority of students have a high degree of consistency in the representation of this concept. However, it is worth noting that the accuracy of LLM image recognition is not high, indicating the possibility of “collective bias” in students'

representation. We found a possible reason for the high consistency rate but low accuracy rate when rechecking children's drawings: some students did not annotate any explanation (only drew three circles representing the sun, earth, and moon, without any signs), and the drawings did not look like them, which led LLM to mistake them for bicycle chains or other unrelated objects.

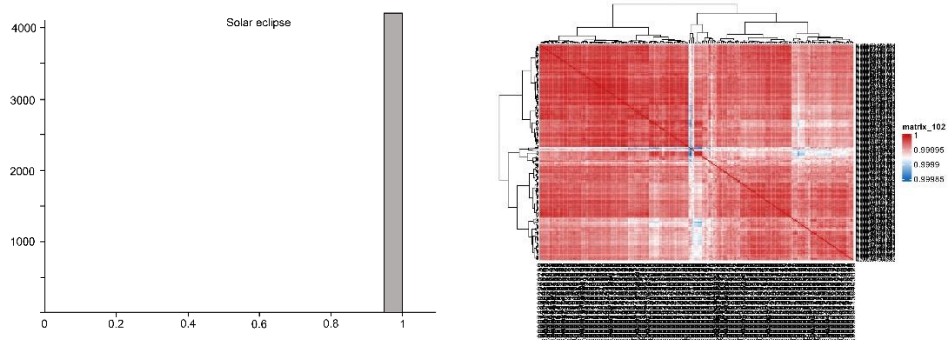


Figure 3: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Solar eclipse"

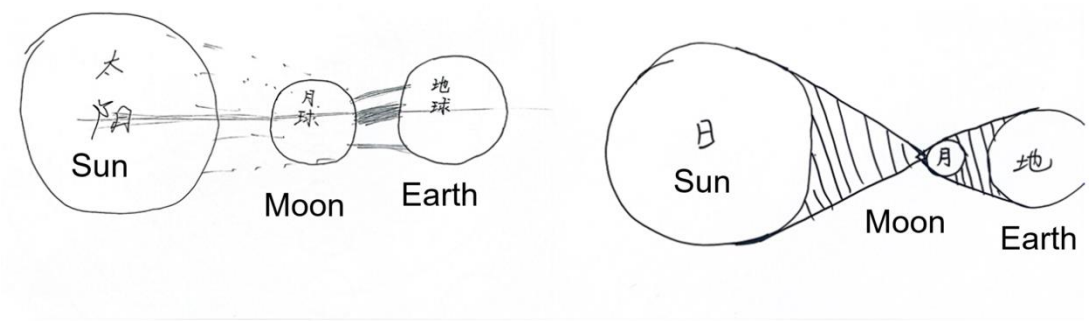


Figure 4 Representative drawing of the concept of "Solar eclipse"

3.2.3 Boiling

The main content (experiment) of this concept is to boil water through alcohol lamp. The textbook requires students to know the phenomenon of water boiling and the temperature changes during this process (i.e., it keeps rising until about 100 ℃ and no longer changes).

The average accuracy of LLM's image recognition is 0, ranking at the bottom of

the nine experiments, indicating that all students' representations misled LLM.

The median value of semantic similarity is 0.967, and 75% Region is concentrated within the 1.49% range of [0, 1]. Although the 75% Region belongs to the second to last of the 9 concepts, the concentration is still very high, indicating that the vast majority of students have a high degree of consistency in the representation of boiling forms. However, the image recognition results of LLM are completely unrelated to this concept, indicating that almost all student representations have problems. This result is very surprising, as can be seen from the representative drawing in Figure 6 where the problem lies: the vast majority of students' drawings focus on the experimental process and equipment, with almost no description of the changes in water during this process. This experiment does indeed require the use of these experimental equipment, but these devices are not what this class is going to learn, which means that students' focus is entirely on the experimental activity itself - rather than what the experiment is going to explain (changes in water), which is truly unexpecting.

It should be noted that we modified a keyword while tuning the model training related to this concept. Due to the significant difference between the results provided by the TF-IDF and our feeling, we conducted a retrospective analysis of all processes and found that an important word, "chemistry," was not recognized as a keyword by TF-IDF. However, this word appeared very frequently in the reasons given by LLM (about 49% of reason sentences mentioned this word). Since the TF-IDF algorithm is a statistical method based on word frequency, it cannot truly understand the meaning of the text, and the word "chemistry" appears only once in each sentence, far less frequently than "this", so it is not surprising that it is excluded by algorithm. Based on this, we selected the function word with the lowest weight from the 10 keywords

provided by TF-IDF, replaced it with "chemistry", modified the corresponding word weight, and attempted to retrain. The results were very close to our intuitive feeling.

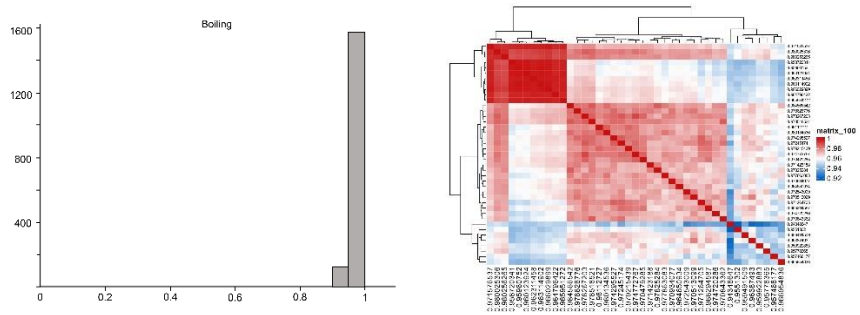


Figure 5: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Boiling"

Figure 6 Representative drawing of the concept of "boiling"

3.2.4 Solar system

The main content of this concept is to understand the solar system, especially the Sun and the eight major planets and their relative positions, but it does not require students learn the motion state and trajectory.

The average accuracy of LLM image recognition is 0.65, which is higher than the median value, indicating that most students can correctly represent this concept.

The median value of semantic similarity is 0.9999, and 75% Region is

concentrated within the 0.001% range of [0, 1]. Similar to the concept 'Circuits', high accuracy and consistency in representation indicate that children have a correct and consistent understanding of this concept.

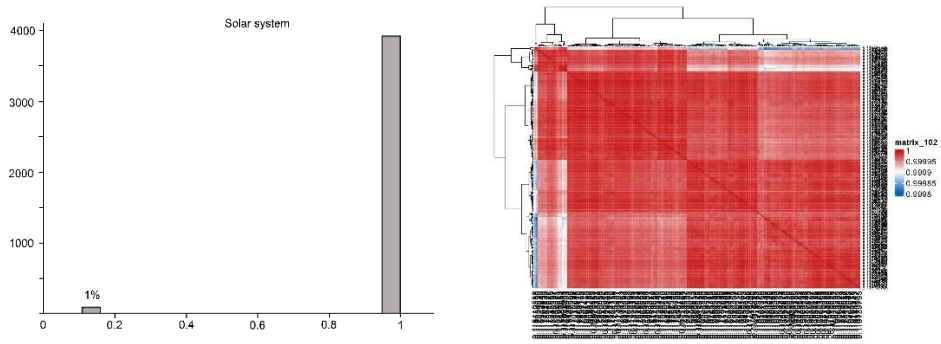


Figure 7: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Solar system", where the percentage in the left figure represents the percentage of data with semantic similarity less than 0.8 in the overall data (the same below)

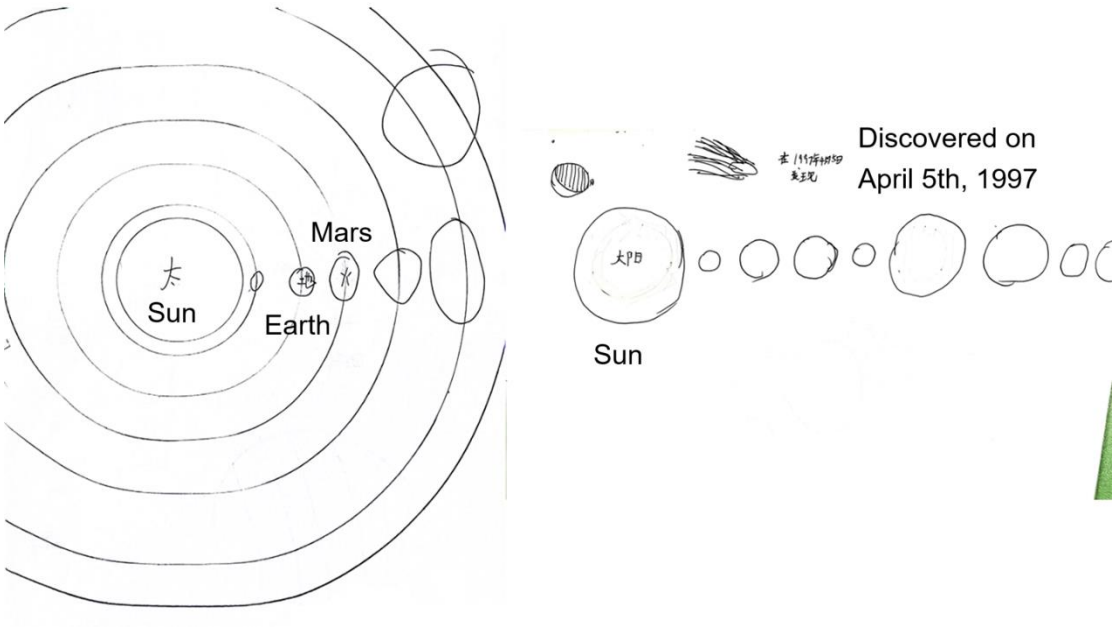


Figure 8 Representative drawing of the concept of "Solar system"

3.2.5 Increasing the carrying capacity of a boat

The main content of this theme in textbook is understanding the relationship between buoyancy and the volume of displacement by changing the volume

and load capacity of the boat. But the presentation of this lesson in the textbook is quite unique: first, boats with different bottom areas were made using aluminum foil of the same area, and then small iron hoops were added to different boats to increase their weight and test their load-bearing capacity. Ultimately, guide students to identify factors associated with buoyancy. That is to say, students cannot directly adjusted variables, but need to first created a 'boat'.

The average accuracy of LLM's image recognition is 0.033, second only to "boiling" and ranking second to last among the nine concepts, indicating that the vast majority of students' representations have misled LLM.

The median value of semantic similarity is 0.987, and 75% Region is concentrated within the 1.29% range of [0, 1]. Overall, the result of this concept is very similar to "Boiling": with low accuracy but high semantic concentration, indicating that children's drawing misled LLM in the same representation way.

However, the reason for this result differs greatly from the concept of "Boiling", which is due to students' excessive focus on the experimental process rather than the experimental purpose, and the problems with this theme are more caused by the textbook. As mentioned above, the original goal of this lesson was to understand the relationship between buoyancy and the volume of displacement. However, in terms of presentation, the textbook first requires students to use aluminum foil to make a boat, and then complete the task of adjusting the weight. That is to say, there is a 'plot transition' in this lesson, and from the representative drawing below, it can be seen that the vast majority of children were stuck at this transition point - they were most impressed with how to make a boat, rather than adjusting variables. Therefore, the most frequently depicted object in the drawing was a square aluminum foil paper with creases

left by the boat making around it (Figure 10). This directly led to errors in the overall direction of LLM: from the perspective of word frequency statistics, 59.6% of LLM's image recognition reasons showed that it was a square box like object that had nothing to do with the boat.

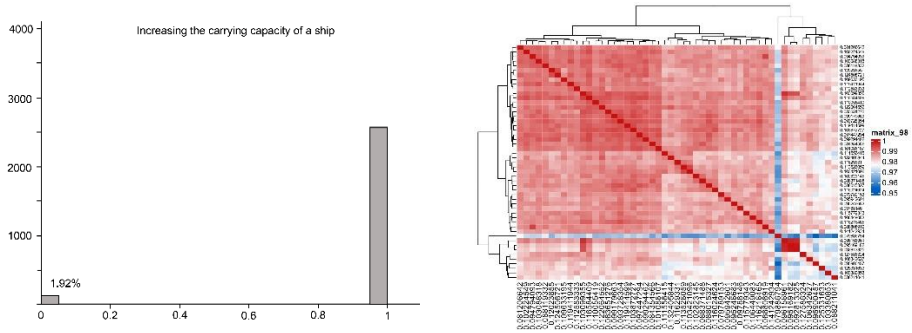


Figure 9: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of " Increasing the carrying capacity of the boat "

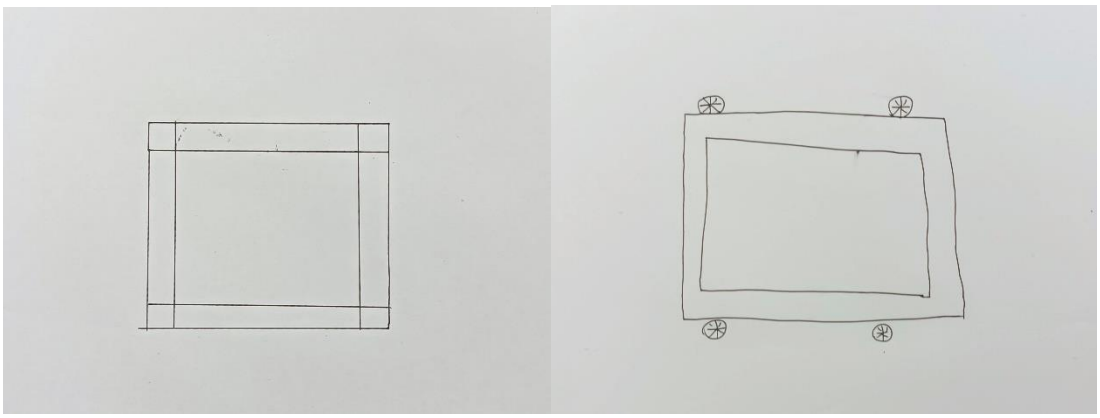


Figure 10: Representative drawing of the concept of "Increasing the carrying capacity of a boat" (the four circular objects in the drawing on the right were small iron hoops used to increase the weight)

3.2.6 Life history of a plant

The main content of this theme in the textbook is the life history of a plant, which refers to the entire process of the plant from seeds to flowering and fruiting. In the previous classes, students have already gone through tasks such as cultivating and caring for plants, and recorded the growth and changes of the plants. Therefore, the main purpose of this lesson is to establish a comprehensive impression of the entire process of the plant development and

changes through previous records.

The average accuracy of LLM image recognition is 0.712, ranking second among the six concepts, indicating that most students can correctly represent this theme.

The median value of semantic similarity is 0.9999, and 75% Region is concentrated within the 0.001% range of [0, 1]. Explain that children have a correct and consistent understanding of this concept.

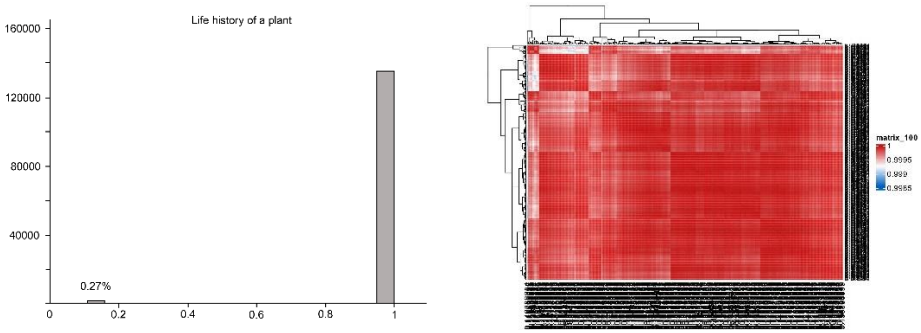


Figure 11: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Life history of a plant"

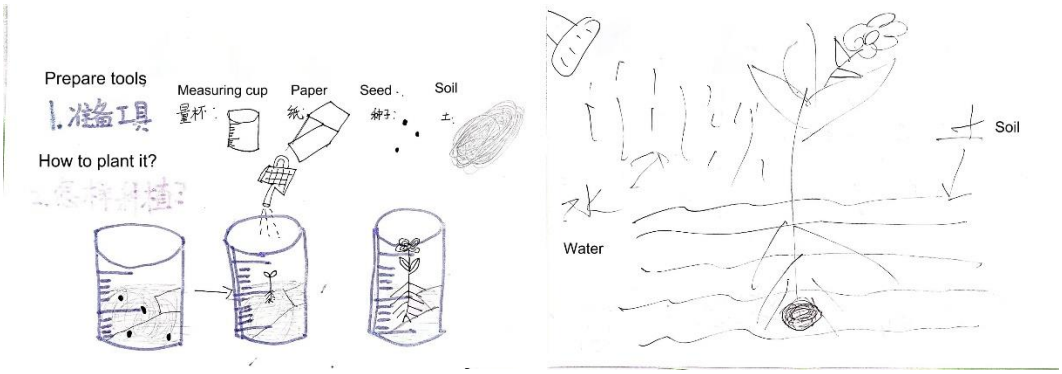


Figure 12: Representative drawing of the concept of "Life history of a plant"

3.2.7 Buoyancy

This theme is very similar to the goal of "increasing the carrying capacity of a boat", which is to control the sinking or floating of the boat by changing its weight and shape. These two themes come from two different versions of textbooks, so there is some overlap in content. But unlike 'increasing the

carrying capacity of a boat', this lesson did not ask students to make a boat, but instead posed the core question at the beginning: how to keep a small boat floating on the water sink? Therefore, when learning this theme, the entire class was focused on exploring how to adjust variables and control the buoyancy of the boat, not to make a boat. So, as will be seen in the following text, there is a significant difference between the results than "Increasing the carrying capacity of the boat".

The average accuracy of LLM's image recognition is 0.5, ranking six among 9 concepts, indicating that a considerable number of students' drawings misled LLM, but it is significantly better than "Increasing the carrying capacity of the boat".

The median value of semantic similarity is 0.944, with 75% Region concentrated within [0, 1] of 3.38%, ranking last among all 9 concepts. After rechecking all the drawings, we found the main reason for the divergence of semantic similarity: the textbook of this lesson did not strictly define what is a "boat", so as shown in the representative drawing below, some students used foam plastic and leather balls. We also found such as plastic balls, wood blocks, iron boxes, well-made submarine models and warship models were used in some drawings. It is precisely because the textbook does not impose too many restrictions that students exhibited rich imagination, which also leads to the diversity of LLM's descriptions of the reasons for image recognition.

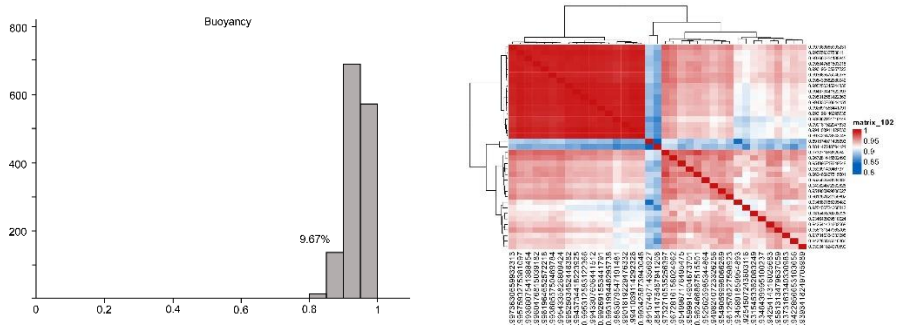


Figure 13: Distribution diagram (left) and heatmap (right) of semantic similarity

of the theme of "Buoyancy"

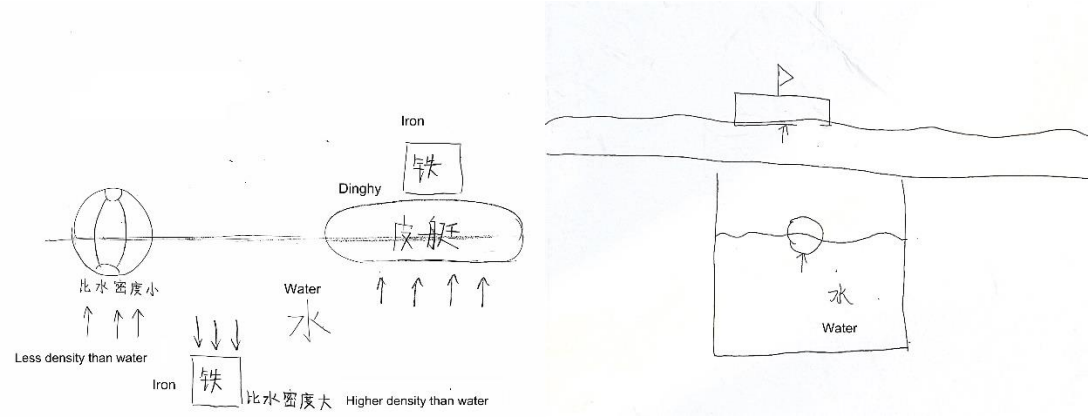


Figure 14 Representative drawing of the theme of “Buoyancy”

3.2.8 Electromagnetics

This concept in textbook is mainly reproduced the experiment of electromagnetic phenomenon discovered by Danish scientist Hans C. Oersted in 1820. And enhance electromagnetic phenomena through wire winding to make coils, to explore related factors that affect the strength of this phenomenon.

The average accuracy of LLM image recognition is 0.514, ranking fourth among 9 concepts.

The median value of semantic similarity is 0.9995, and 75% Region is concentrated within the 0.11% range of [0, 1], indicating that students have strong consistency in their representation of this concept.

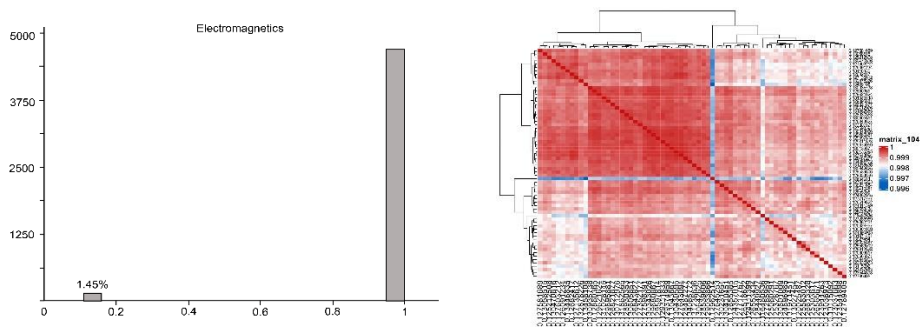


Figure 15: Distribution diagram (left) and heatmap (right) of semantic similarity

of the concept of "Electromagnetics"

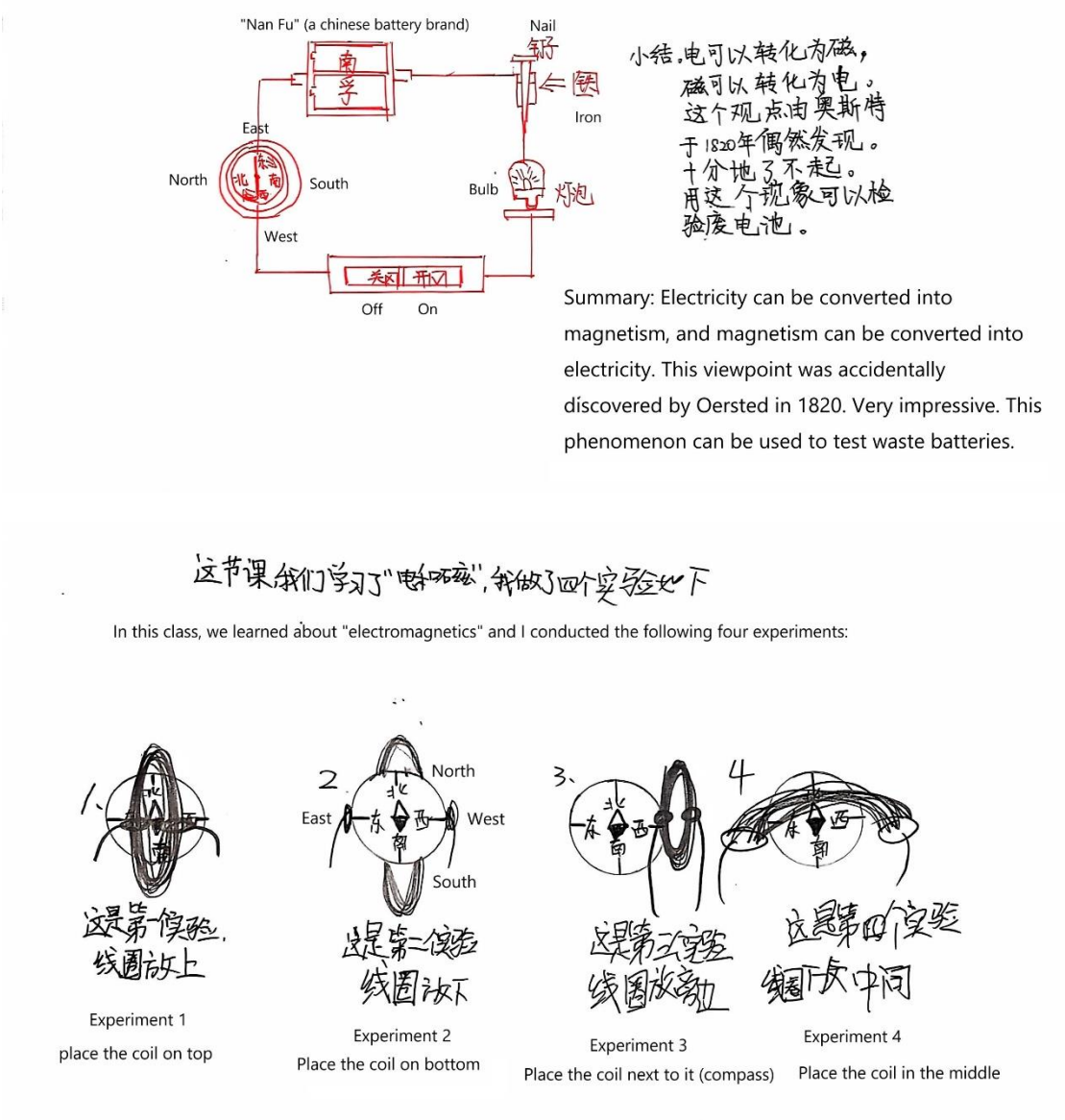


Figure 16 Representative drawing of the concept of "Electromagnetics"

3.2.9 Physical and chemical changes

The main content of this theme is to explore various changes in the kitchen, and to understand the difference between "physical changes" and "chemical changes" based on whether new substances are produced. This theme mainly corresponds to the "Next Generation Science Education Standards" (NGSS): "When two or more different substances are mixed, a new substance with different properties may be formed" (5-PS1-4) [20].

The average accuracy of LLM image recognition is 0.513, ranking fifth among the 9 concepts, that is, in the middle.

The median value of semantic similarity is greater than 0.9999, and 75% Region is concentrated within almost 0% range of [0, 1], indicating that students have a huge strong consistency in their representation of this theme. It should be noted that, as shown in the representative drawing below, children exhibit two completely different tendencies towards drawing this theme: one is to express it through graphics, and the other is to express it through mind maps. However, from the perspective of semantic similarity, the difference between these two forms of expression does not affect the understanding of semantics by the word2vec algorithm, because even when using graphics to represent this theme, a considerable number of students have added prompts in the drawings. The detailed information will be elaborated in the "Discussion" section of this article.

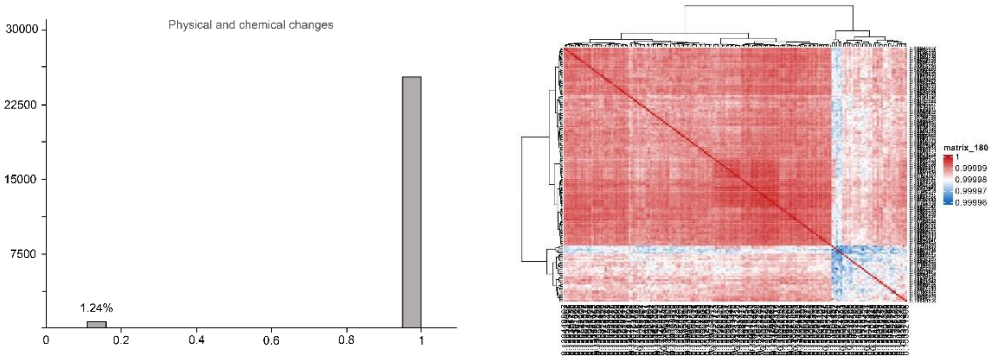
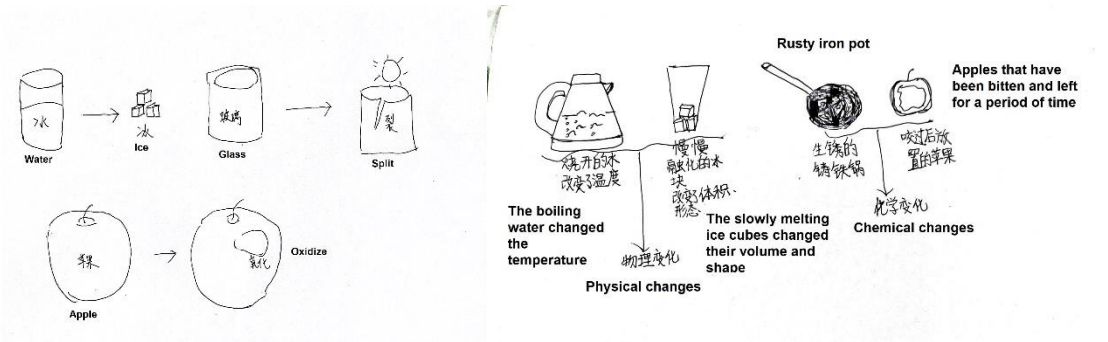


Figure 17: Distribution diagram (left) and heatmap (right) of semantic similarity of the concept of "Physical and chemical changes"



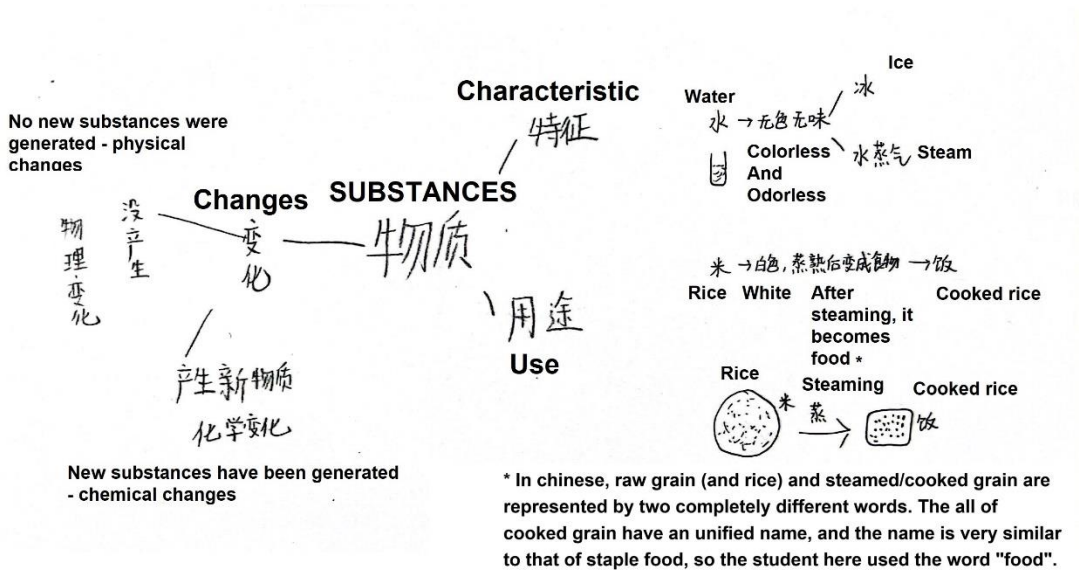


Figure 18: Representative drawing of the theme of "Physical and chemical changes"

3.3 Accuracy of Human Recognition

Due to the fact that the above results were generated by LLM's recognition, in order to ensure the reliability of the results, we recruited 45 adults to recognize 27 drawings (3 randomly selected drawings for each theme), and analyzed their correctness and reasons given. The reason why we only selected 27 drawings is because once the number of drawings on the same theme exceeds three, the participants will guess the theme through previous drawings, this is something LLM cannot do.

In terms of accuracy, 85.19% of the themes of drawings that humans and machines made roughly consistent judgments, indicating that the machine's results in terms of image recognition accuracy are reliable. However, in terms of semantic similarity, the average median value of the nine concepts of human similarity is only 0.372, which is far from the results of LLM. Due to the lack of further evidence, we are unable to explain this result. However, there is a logical possibility that human language generation ability is much stronger than that of

machines, and therefore, the diversity of reasons for image recognition is much more complex than LLM, resulting in a significant difference in semantic similarity between machines and humans.

3.4 Analysis of Related Factors

As mentioned above, this article uses Kendall rank correlation coefficient (Kendall-τ) to analyze the factors that affect drawing's representations.

3.4.1 Relationship with sample sizes

Is there a relationship between the above results and the total sample size? We separately calculated the correlation coefficients between "sample size" and three factors ("accuracy", "median value of semantic similarity", and "75% Region"), and the results showed that the correlation coefficients of the three factors were 0.611, 0.609, and -0.493, respectively (the three bar charts above Figure 19). It can be seen that sample size has a strong correlation with accuracy and median value of semantic similarity, and has a moderate correlation with semantic similarity concentration, and this moderate correlation is not significant (p=0.07).

3.4.2 Relationship with the abstract degree of themes/concepts

Based on J. Piaget's theory of developmental stages in children, the abstract thinking ability of elementary school students has not yet been fully developed [21]. Therefore, we speculate whether the above results are related to the degree of abstraction of concepts? That is to say, the higher the degree of abstraction, the lower the accuracy and the lower the semantic similarity? However, how to define the degree of abstraction of concepts? Children and adults have different understandings of abstraction. Some adults believe that concrete concepts (such as mathematical laws of force or motion) are abstract and difficult to grasp in the eyes of children because they cannot be directly

seen [22].

For this purpose, we coded the abstract degree of all 9 themes/concepts. The coding rule is: a phenomenon or something that can be seen or directly felt by sensory organs without any intervention was coded 1, otherwise it was coded 2. Based on this rule, the following encoding is obtained:

Themes/concepts	Codes	Reasons of codes
Circuit	1	A shape (the focus of this lesson is to make students realize that bulb can be lighted only when the circuit has a “closed circular” shape)
Solar eclipses	1	A phenomenon
Boiling	1	A phenomenon
Solar system	1	A phenomenon
Increasing the carrying capacity of a boat	2	It involves the regulation of variables and cannot be directly seen or felt
Life history of a plant	2	Different plants have different growth processes, with significant differences
Buoyancy	2	The content of this lesson is not about explaining buoyancy itself, but about adjusting the magnitude of buoyancy, involving the regulation of variables
Electromagnetics	1	A phenomenon
Physical and chemical changes	2	The scope of these two types of changes is broad and can be represented by multiple exemplars

Table 2: The codes of abstract degree and specific reasons for 9 concepts

We invited 37 science teachers (excluding researchers) to evaluate the above coding rule, of which 31 fully agreed (83.8%); 4 out of the remaining 6 people do not agree with “Life history of a plant”, mainly because the textbook uses Impatiens as the main exemplar to show this theme, and there are fewer other plants involved; 3 people do not agree with “Buoyancy”, mainly because the lesson also includes the content of feeling buoyancy by pressing a small bottle, so it is not only the regulation of variables; One person simultaneously questioned both of the above themes. So overall, the above coding rule have been recognized by most science teachers.

As mentioned above, we still use Kendall-tau to analyze the correlation between abstract degree and "accuracy", "median value of semantic similarity", "75% Region". The results showed that the correlation coefficients of the three factors were -0.314, -0.301, and 0.246, respectively (the three bar charts in the middle of Figure 19), indicating that the abstract degree of the concept is not significantly related to the above three factors.

3.4.3 Reproducibility the contents in the classroom

Another factor we are concerned about is whether students will express abstract themes/concepts by reproducing contents that have appeared in class through drawings? Due to the elusive of these themes/concepts for children, textbooks often present them through representative examples, such as express the “Life history of a plant” through Impatiens. So, will students represent these difficult things by reproducing the examples that have appeared in the classroom? We used the examples that appeared in the textbook as the basis and calculated the proportion of these example words in the LLM’s recognition reasons through word frequency statistics. And calculated the total proportion using weighted average (weighted by sample size). The results are as follows:

Themes/concepts	**Sample size**	**Count of reproduction**	**Rate of reproduction %**	**Examples ****
Increasing the carrying capacity of a boat	52	24	46.15	Square, rectangle
Life history of a plant	368	—— *	——	——
Buoyancy	40	8	20	Bottle, submarine
Physical and chemical changes	70	160	100	Water, salt, flour, dough, iron pot, rust, eggs, apples, ice cubes

* It is truly difficult for us to identify whether the plants in drawings were Impatiens or not (most of the drawings do not provide the names of the plants), so we did not calculate the data for this theme.

** Due to the large number of examples appearing in this lesson, it is not required that all examples be drawn. As long as one example appears in the drawings, it will be counted.

Table 3 Reproduction rate of samples encountered in the classroom

The weighted average reproduction rate is 76.19%, which means that at least three-quarters of the students have reproduced the classroom content in the drawings more or less to represent these difficult things. Considering that LLM's recognition accuracy is not 100%, this proportion is likely to be even higher.

3.4.4 Relationship with the focus points

This is what we discovered after carefully examining the drawings: we found that the concept of "boiling" originally required students to understand the process of water to steam change, but most of the students drew the experimental activities (i.e. the experimental apparatus). This experiment is an indispensable process for heating water, but not the purpose of the lesson. However, it was used by students as a representation of the phenomenon of water boiling. In other words, there is a significant deviation between the students' focus and the purpose in the textbook.

Therefore, we developed a coding rule based on the role of experiments in lessons, and used it as an independent variable to conduct correlation analysis on the above three factors. The coding rule are as follows: those themes without experiments or experimental results were equivalent to the teaching purposes will be coded 1; The experimental results were not the teaching purposes, but only the indispensable process of teaching purposes will be coded 2 (simple

drawing and hands activities are not considered experiments). The coding result is as follows:

Themes/concepts	Codes	Reasons of codes
Circuit	1	The experimental results are consistent with the teaching purpose, both pointing to the shape of the circuit (a "circle")
Solar eclipses	1	There is no experiment in this class
Boiling	2	The experiment is a necessary condition for boiling, but it is not the teaching purpose (i.e. it is not the boiling phenomena and changes in water)
Solar system	1	There is no experiment in this class
Increasing the carrying capacity of a boat	2	This lesson contains two activities, the first of which is not related to the teaching purpose and only about creating a small boat that can boat in the water
Life history of a plant	1	There is no experiment, just a summary of previous plant planting activities
Buoyancy	1	The experimental results are consistent with the teaching purpose, both pointing to regulating the sinking and floating of the small boat
Electromagnetics	1	The experimental results are consistent with the teaching purpose, both pointing to electromagnetic phenomena
Physical and chemical changes	1	The experimental results are consistent with the teaching purpose, both pointing to demonstrating a physical or chemical change

Table 4: Coding and Reasons for 9 Concepts about the relationship between experimental results and teaching purpose

Due to the fact that the above coding rule involve a deep understanding of textbook content, we invited two senior science teachers (both of whom have senior professional titles) to review the above rules and received their unanimous agreement.

The correlation coefficients are -0.624, -0.488, and 0.465 (the three bar charts below Figure 19), indicating that the consistency between experimental results and teaching purposes can explain the accuracy very well. Specifically, it can effectively solve the problem of "collective bias" in students' drawings: students are more concerned with the experiment itself, but if the experiment deviates

from the teaching purposes, students are likely to choose to remember the experiment rather than the real purposes. Meanwhile, there is a moderate correlation with the other two items (median value of semantic similarity value and 75% Region).

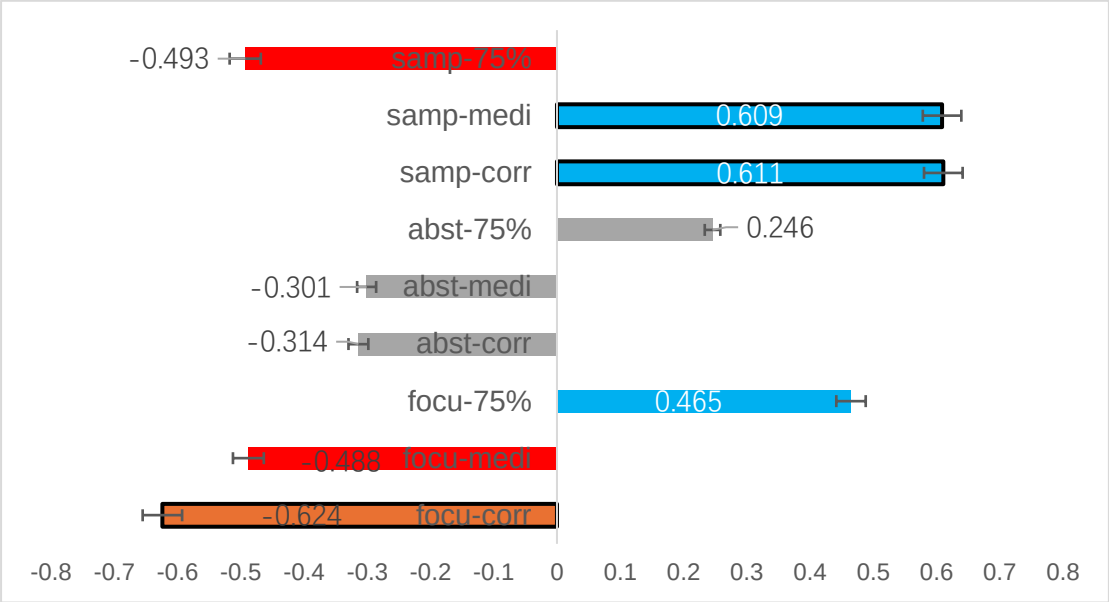


Figure 19: shows the Kendall rank correlation coefficient results for each factor, where “samp” represents the factor of sample size, “abst” represents the factor of abstract degrees, “focu” represents the factor of focus points, , "75%" represents 75% Region, “medi” represents median, and “corr” represents accuracy; “samp-medi” represents the correlation coefficient between the sample size and the median value (other similar); Color represents moderate or above correlation (tau>0.4), and bold black boxes represent the correlation is significant ($p<0.05$)

4. Discussion

This article outlines the representations of 9 scientific concepts in children's eyes through 1420 children's scientific drawings, and explores whether there is a consistent representation in children's scientific drawings through LLM image recognition and semantic similarity analysis. The results show that the representation of most drawings has consistency, manifested as most semantic

similarity>0.8. At the same time, it was found that the consistency of the representation is independent of the accuracy (of LLM's recognition), indicating the existence of consistency bias. In the subsequent exploration of influencing factors, we used Kendall rank correlation coefficient to investigate the effects of "sample size", "abstract degree", and "focus points" on drawings, and used word frequency statistics to explore whether children represented abstract themes/concepts by reproducing what was taught in class. It was found that accuracy (of LLM's recognition) is the most sensitive indicator, and data such as sample size and semantic similarity are related to it; The consistency between classroom experiments and teaching purpose is also an important factor, many students focus more on the experiments themselves rather than what they explain. In addition, most children tend to use examples they have seen in class to represent more abstract themes/concepts, indicating that they may need concrete examples to understand abstract things.

4.1 The contradiction between accuracy and consistency of representation

However, looking back at the entire result, we found a rather strange thing: some themes/concepts (solar eclipse, buoyancy, electromagnetics, physical and chemical changes) have an accuracy rate close to 50%, but their semantics are very close to each other (median>0.99). In this study, semantic similarity is used to represent the consistency of students' drawing representations. A high accuracy rate and high semantic similarity indicate that students are able to correctly and consistently represent the concept with drawings. A low accuracy rate but high semantic similarity indicates that students may have a problem of "consistency bias". So, how can we explain why the accuracy is close to 50% but the semantic similarity is high? The accuracy rate approaching 50% indicates that there should be significant differences in students' drawings (i.e., the number of correct and incorrect drawings is almost equal), why is the semantic similarity still high? This seems somewhat abnormal.

We carefully reviewed these drawings and found a possible "suspect": some drawings that LLM deemed incorrect actually used representations that were very similar to those of the correct drawings, but lacked some key prompts, thus misleading LLM. As shown in Figure 20, both the left and right images depict the transformation of water into ice and the oxidation of apples after being bitten, but the difference is that the left image clearly indicates "physical changes" or "chemical changes" below each change, while the right image does not indicate these words. So, the left image was correctly recognized by LLM, but the right image was not.

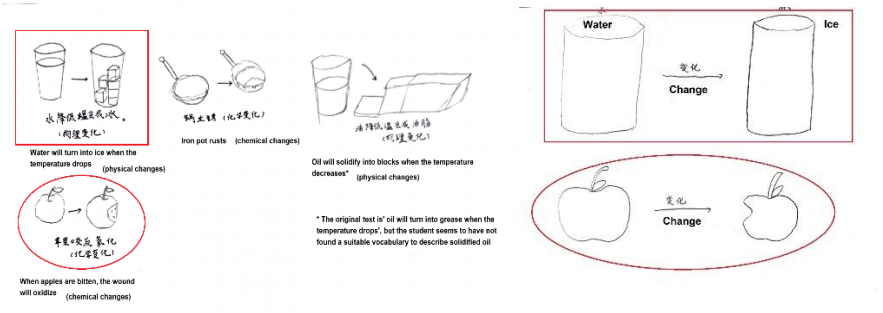


Figure 20 Both left and right images depict the process of water turning into ice and the oxidation of apples

4.2 About Abstract Concepts

Another issue we are concerned about is how children understand abstract concepts, because obviously, abstract things are more difficult for people to learn concepts than concrete concepts, especially for children. For example, Elika Bergelson et al. found that infants and young children have a significantly later understanding of abstract vocabulary than concrete vocabulary [23], but learning abstract things is an important task in education.

Previous research has focused on the following directions: firstly, utilizing language itself. For example, Gabriella Vigliocco et al. proposed after reviewing previous research that children can use context to assist in understanding more

abstract concepts [24]. In fact, this is also the basic principle based on which mainstream NLP algorithm models are currently based [25]. In addition, Yanchao Bi et al. found based on fMRI that semantic processing is also more involved in the cognition of abstract concepts of human [26]. Secondly, using valence to understand [24], adults also use this effect to understand abstract concepts [27]. The third factor is the involvement of visual processing, as Amedeo D'Angiuli et al. found that as abstract words are deeply processed, occipital lobe activity is significantly activated, indicating that children are likely to begin "imagining" the word [28].

However, for drawings, it is not entirely the same, as this approach clearly requires children to react and express their understanding of abstract concepts (i.e., they need to demonstrate it through behaviors), which is significantly different from the results obtained by directly scanning the brain using neuroscience equipment mentioned earlier. The previous review showed that there are two tendencies in children's drawing towards abstract concepts: one is influenced by language, and the other is to search for prototypes that can represent this abstract concept (which seems to be similar to the visual processing mentioned earlier, as finding prototypes is a concrete process that clearly requires imagination [28]) [22].

We also found these two clear tendencies in this study, and we found that among the 9 themes/concepts we explored, one theme clearly belongs to the abstract category: Physical and chemical changes (obviously, these two terms represent two major types of changes, rather than two concrete concepts). Therefore, the following analysis is mainly focused on this theme.

Firstly, as mentioned earlier, regarding the drawing of this theme, children have more or less reproduced the experiments in the classroom 100%. It can be seen

that they have signs of searching for prototypes, because obviously the experiments once appeared in the classroom, which can naturally become typical exemplars for establishing prototypes of "Physical and chemical changes".

Secondly, the tendency to use language to represent this abstract theme is also very evident. Data shows that in 160 children's drawings, 53.8% of them contain at least one of prompts of "physical changes" and "chemical changes", and there are also many drawings that contain both. In addition, 89.5% of these drawings intentionally divide physical or chemical changes into two groups (left in Figure 21); More representative is that 13.1% of the drawings do not have any graphics at all, but instead fully utilize mind maps to present this theme (right in Figure 21). This seems to reflect that when students find it difficult to represent a concept with concrete graphics, they tend to use textual explanations as a concise and intuitive way to solve this problem. The data for judging the drawings mentioned above comes from two researchers who independently assessed three indicators of the drawings: Whether the drawing contains prompts for "physical changes" or "chemical changes"? Is the drawing intentionally grouping examples of the two changes? Is this drawing a graphic or a mind map? The Kappa coefficient is equal to 0.93 (95% CI [0.88, 0.96]), indicating a very high consistency rate between the two individuals' judgments and a high level of credibility in the results.

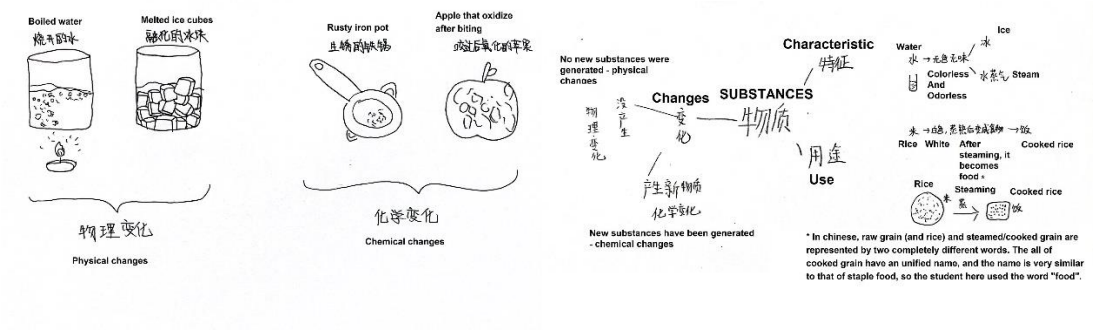


Figure 21 Left: using pictures to represent this abstract theme, and dividing the pictures into two groups -physical changes (left) and chemical changes

(right); Right: Using mind maps to represent this abstract theme

4.3 About the focus points

What we are most concerned about is undoubtedly the factor of "focus points", because the results presented in this article do not seem to be completely consistent with the ideas advocated by international science education at present.

Undoubtedly, in the field of science education today, it is the main steam to advocated to teach more implicit abilities such as reasoning and critical thinking in science classrooms, rather than just allowing students to memorize and apply concrete knowledge. Both Bloom's theory [29] and internationally renowned literature on science education such as NGSS [30] or PISA [31] place great emphasis on these implicit abilities. In other words, the lessons advocated by science education should not only focus on knowledge and inquiry activities themselves, but should also encourage students to infer the true results from these specific activities - that is, "what these experimental results indicate/prove".

However, children do not seem to think so. From the experimental results, it can also be seen that their focus is not the same as that advocated by adults and science education. They are more sensitive to concrete and vivid things, and have a deeper impression. In fact, this result shifts the focus to another question: Who is the advantage, visual processing or semantic processing in children's cognition? Because obviously, implicit cognitive activities such as reasoning and proving conclusions through experimental results require a significant amount of semantic processing. But if one student has a deeper impression of the experimental process or equipment, it is obviously more due

to visual processing. In this regard, the results clearly support the winner is visual processing. Many studies have confirmed that more complex semantic processing can only occur in older children [32] [33], and there is also some evidence that children have a deeper memory of familiar things, indicating that the coupling relationship between vision and memory may be closer [34].

Perhaps readers may think that a shift in focus points can be achieved through classroom training, but we believe that this shift is not very easy to achieve because, as revealed by the experimental results of this article, the divergence of focus seems to be a common law among children, which has a huge inertia. The literature listed above also shows that these inertias have fundamental support at the level of neuroscience. Therefore, when writing textbooks and setting teaching purposes, we should not only rely on curriculum ideas, but also pay attention to the cognitive development laws of children. What is the appropriate degree of implicit components in a lesson? How far is the distance between specific activities and implicit teaching purposes for students can reach? These all require careful consideration and demonstration.

Of course, due to the limitations of researchers' own abilities, the number of drawings we can process is not yet sufficient to establish a norm that covers all themes/concepts in science education. But we believe that the above results can provide some reference value for children's drawing research and reveal some cognitive rules of children towards scientific concepts. We also hopes that more researchers can further develop the methods in this article to analyze more children's drawings, in order to truly establish a norm that covers all themes/concepts.

5. Data accessibility:

5.1 All experimental data has been uploaded to: https://www.scidb.cn/s/yel3ay

5.2 Due to the involvement of children's personal information, if need to view the original drawings, please contact the corresponding author of this article to obtain electronic photos of the original drawings.

Reference:

[1] Fabris MA, Lange-Küttner C, Shiakou M and Longobardi C (2023) Editorial: Children's drawings: evidence-based research and practice. *Front. Psychol.* 14:1250556. https://doi.org/10.3389/fpsyg.2023.1250556

[2] P. Bozzato, Matteo Angelo Fabris &C. Longobardi. Gender, stereotypes and grade level in the draw-a-scientist test in Italian schoolchildren. *International Journal of Science Education.* 2021.10, 2640-2662, https://doi.org/10.1080/09500693.2021.1982062

[3] Brechet C, Blanc N, Mortier A and Rossi S (2022) Draw me a brain: The use of drawing as a tool to examine children's developing knowledge about the “black box”. *Front. Psychol.* 13:951784. https://doi.org/10.3389/fpsyg.2022.951784

[4] Marengo D, Settanni M, Longobardi C and Fabris MA (2022) The representation of bullying in Italian primary school children: A mixed-method study comparing drawing and interview data and their association with self-report involvement in bullying events. *Front. Psychol.* 13:862711. https://doi.org/10.3389/fpsyg.2022.862711

[5] Tolsberg K, Põldre S and Kikas E (2022) Learner-generated drawing as a learning strategy. The effect of teacher-guided intervention program “Learning with Understanding” on composing drawings in math word problems in the primary grades. *Front. Educ.* 7:962067. https://doi.org/10.3389/feduc.2022.962067

[6] Restrepo Cervantes, D., Chamorro Coneo, A., Bolivar Pimiento, D. et al. A Psychometric Analysis of the Moral Disengagement Scale (MDS) in Association to Bullying Roles in Colombian Youth. *Int Journal of Bullying Prevention* (2024). https://doi.org/10.1007/s42380-024-00215-y

[7] Marengo D, Settanni M, Longobardi C and Fabris MA (2022) The representation of

bullying in Italian primary school children: A mixed-method study comparing drawing and interview data and their association with self-report involvement in bullying events. *Front. Psychol.* 13:862711. https://doi.org/10.3389/fpsyg.2022.862711

[8] *Front. Psychol.* Children's Drawings: Evidence-based Research and Practice. https://www.frontiersin.org/research-topics/24933/childrens-drawings-evidence-based-research-and-practice/overview

[9] Beltzung B, Pelé M, Renoult JP and Sueur C (2023) Deep learning for studying drawing behavior: A review. *Front. Psychol.* 14:992541. https://doi.org/10.3389/fpsyg.2023.992541

[10] Bo YU et al. *SCINCE.* Beijing: Educational Science Publishing House. 2017

[11] Wei DUAN, Xiang PENG et al. *SCIENCE.* Changsha: Hunan Science & Technology Press. 2017

[12] Maolin Zhang, Xianyong Li, Yajun Du et al. A Multimodal Large Language Model for Forestry in pest and disease recognition. *Research Square.* https://doi.org/10.21203/rs.3.rs-3444472/v1

[13] Shuo Yang, Zirui Shang, Yongqi Wang et al. Data-free Multi-label Image Recognition via LLM-powered Prompt Tuning. *Arxiv.* https://doi.org/10.48550/arXiv.2403.01209

[14] Baike.baidu.com. Word2vec. https://baike.baidu.com/item/Word2vec?fromModule=lemma_search-box

[15] Yoav Goldberg, Omer Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *Arxiv.* https://doi.org/10.48550/arXiv.1402.3722

[16] Faisal Rahutomo, Teruaki Kitasuka, Masayoshi Aritsugi. Semantic Cosine Similarity. *Conference: The 7th International Student Conference on Advanced Science and Technology ICAST 2012.* https://www.researchgate.net/publication/262525676_Semantic_Cosine_Similarity

[17] Syukry Ansis, Endang Palupi Listyaningsih, Hari Soetanto. DETEKSI PLAGIAT TESIS BERBAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY. INOVTEK, *Polbeng - Seri Informatika,* 2024.06, 09(01), https://doi.org/10.35314/isi.v9i1.4003

[18] J. Lilleberg, Y. Zhu and Y. Zhang, "Support vector machines and Word2vec for text classification with semantic features," *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, Beijing, China, 2015, pp. 136-140, https://doi.org/10.1109/ICCI-CC.2015.7259377

[19] Peter Flom et al. Common errors in statistics and methods. *BMJ Paediatrics Open* 2024; 8: e002755. https://doi.org/10.1136/bmjpo-2024-002755

[20] National Research Council, Zhaoning YE et al. (Trans.). *Next Generation Science Standards: Disciplinary Core Ideas.* Beijing: China Science and Technology Press, 2020, 44

[21] Galotti KM. *Cognitive Psychology: In and Out of the Laboratory*, Beijing: China Machine Press. 2017. 218

[22] Zhong Wang. Learning of Concepts: A Review of Relevant Advances Since 2010 and Its Inspirations for Teaching. *Journal of Contemporary Educational Research.* 2024, 08(06), https://doi.org/10.26689/jcer.v8i6.7049

[23] Elika Bergelson, Daniel Swingley. The acquisition of abstract words by young infants. *Cognition.* 2013, 127(03): 391-397, https://doi.org/10.1016/j.cognition.2013.02.011

[24] Gabriella Vigliocco, Marta Ponari, Courtenay Norbury. Learning and Processing Abstract Words and Concepts: Insights From Typical and Atypical Development. *Topics in cognitive science.* 2018, https://doi.org/10.1111/tops.12347

[25] Ashish Vaswani et al. Attention is all you need. *CoRR,* abs/1706.03762, 2017b. URL, http://arxiv.org/abs/1706.03762

[26] Striem-Amit, E., Wang, X., Bi, Y. et al. Neural representation of visual concepts in people born blind. *Nat Commun* 9, 5250 (2018). https://doi.org/10.1038/s41467-018-07574-3

[27] Xiaosha Wang, Wei Wu, Zhenhua Ling et al. Organizational Principles of Abstract Words in the Human Brain. *Cerebral Cortex,* Volume 28, Issue 12, December 2018, Pages 4305–4318, https://doi.org/10.1093/cercor/bhx283

[28] D'Angiulli A, Griffiths G and Marmolejo-Ramos F (2015) Neural correlates of visualizations of concrete and abstract words in preschool children: a developmental

embodied approach. *Front. Psychol*. 6:856.
https://doi.org/10.3389/fpsyg.2015.00856

[29] Nancy E. Adam. Bloom's taxonomy of cognitive learning objectives. *J Med Libr Assoc*. 2015 Jul; 103(03):152-153, https://doi.org/10.3163/1536-5050.103.3.010

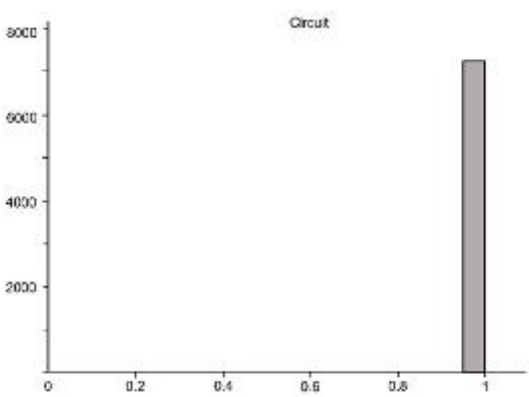
[30] Grooms, Jonathon; Enderle, Patrick; Sampson, Victor. Coordinating Scientific Argumentation and the Next Generation Science Standards through Argument Driven Inquiry. *Science Educator,* 2015, 24(01): 45

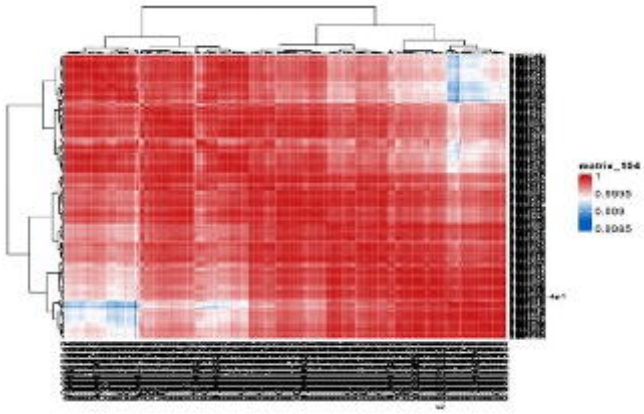
[31] Chun-Yen Tsai. The effect of online argumentation of socio-scientific issues on students' scientific competencies and sustainability attitudes. *Computers & Education,* 2018, 116: 14-27. https://doi.org/10.1016/j.compedu.2017.08.009

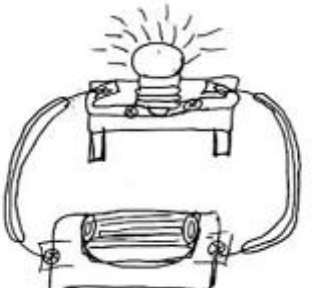
[32] Tai-Li Chou, James Booth, Tali Bitan et al. Developmental and Skill Effects on the Neural Correlates of Semantic Processing to Visually Presented Words. *Human Brain Mapping.* 27(11), 915-924, https://doi.org/10.1002/hbm.20231

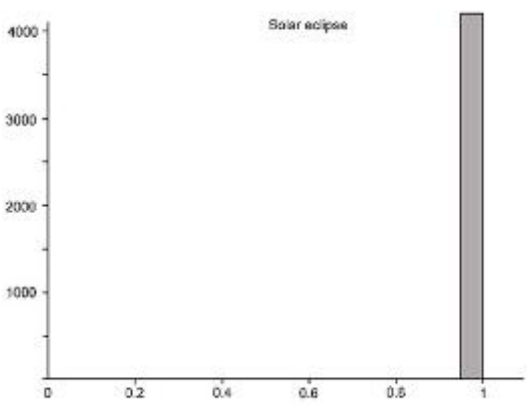
[33] Lucy A. Henry, David Messer. Phonological, visual, and semantic coding strategies and children's short-term picture memory span.
Quarterly Journal of Experimental Psychology. 2012.10,
https://doi.org/10.1080/17470218.2012.672997

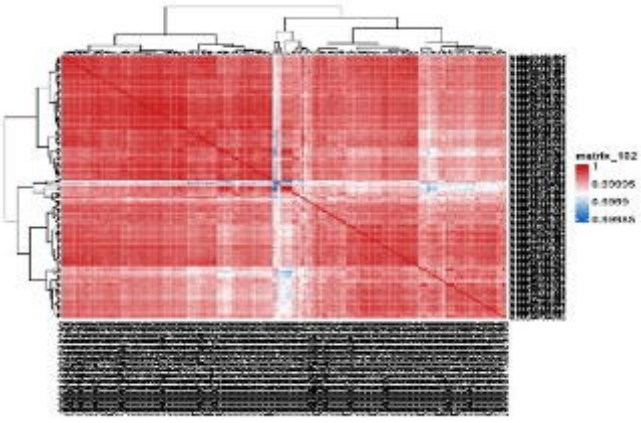
[34] Ariel Starr, Mahesh Srinivasan, Silvia A. Bunge. Semantic knowledge influences visual working memory in adults and children. *PLOS ONE.* 2020.11,
https://doi.org/10.1371/journal.pone.0241110

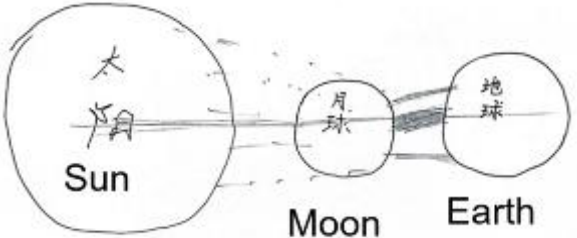

Circuit
8000
6000
4000
2000
0
0.2
0.4
0.6
0.8
1

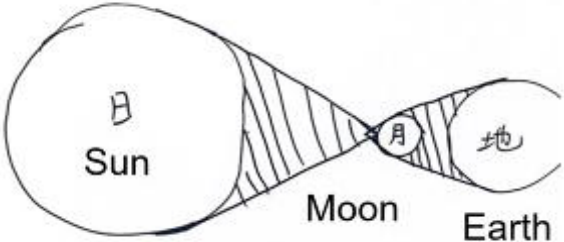


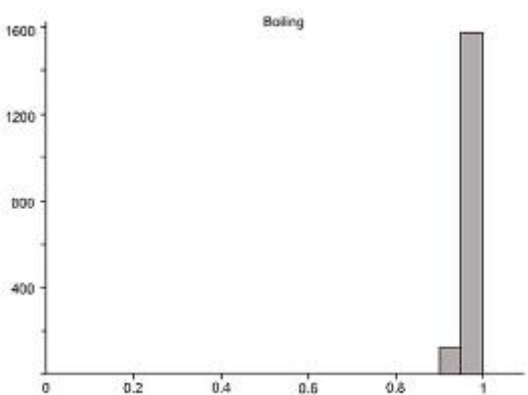


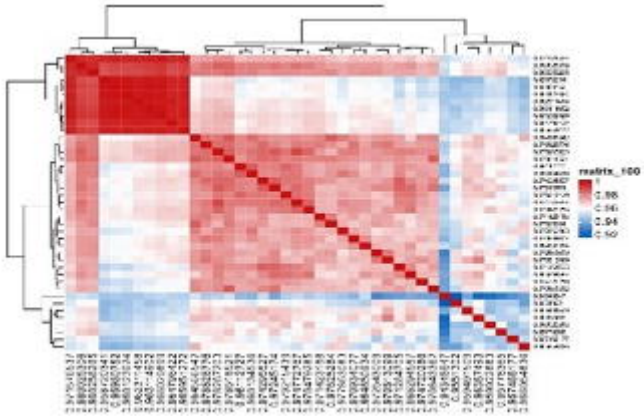

Solar eclipse
4000
3000
2000
1000
0
0.2
0.4
0.6
0.8
1

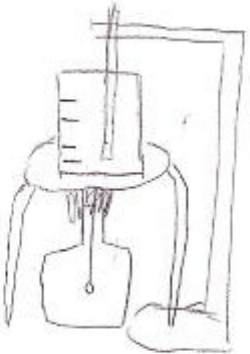


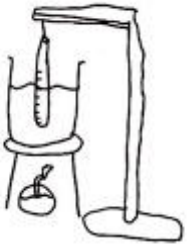

太陽
Sun
月球
Moon
地球
Earth

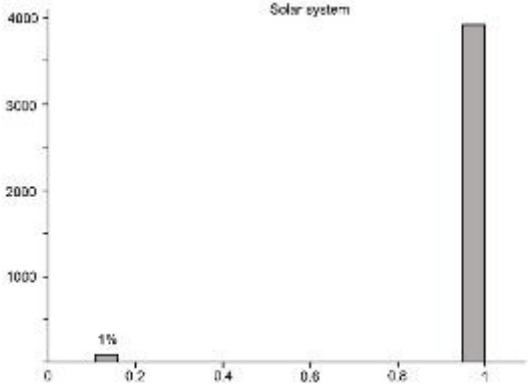

日
Sun
月
Moon
地
Earth

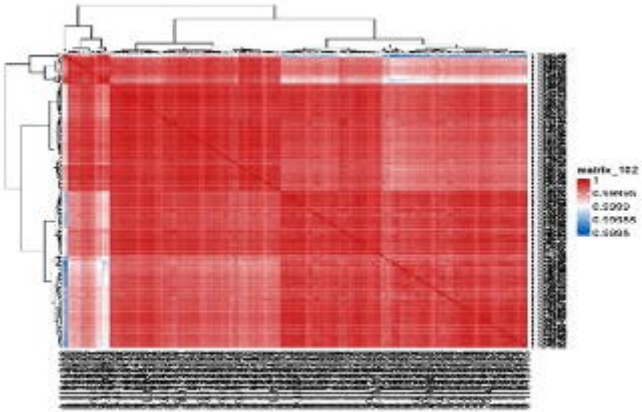

Boiling
1600
1200
800
400
0
0.2
0.4
0.6
0.8
1

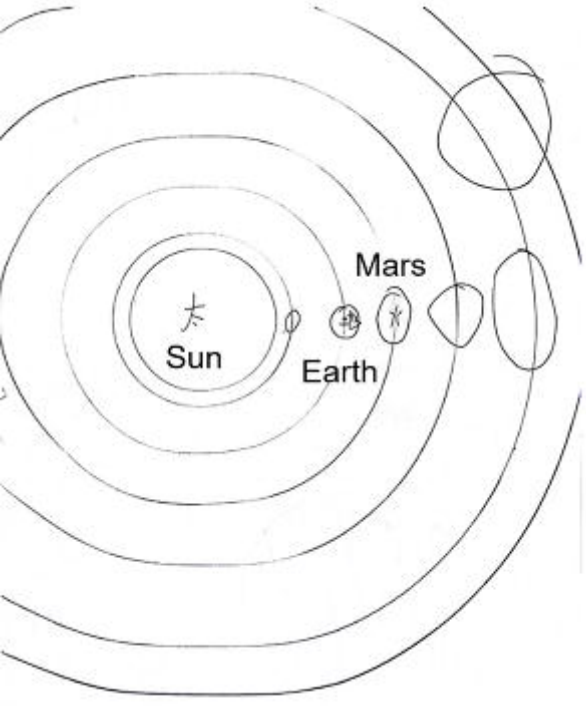


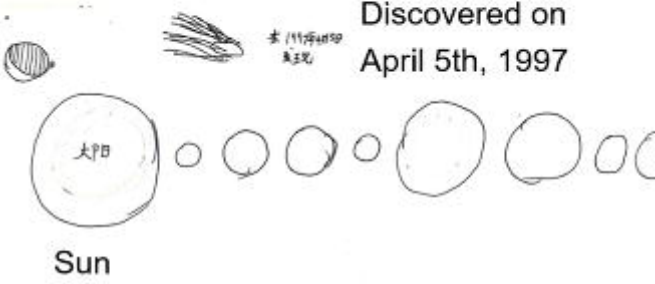




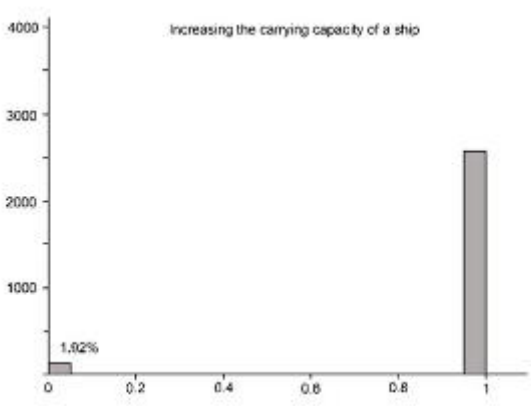

Solar system
4000
3000
2000
1000
1%
0
0.2
0.4
0.6
0.8
1

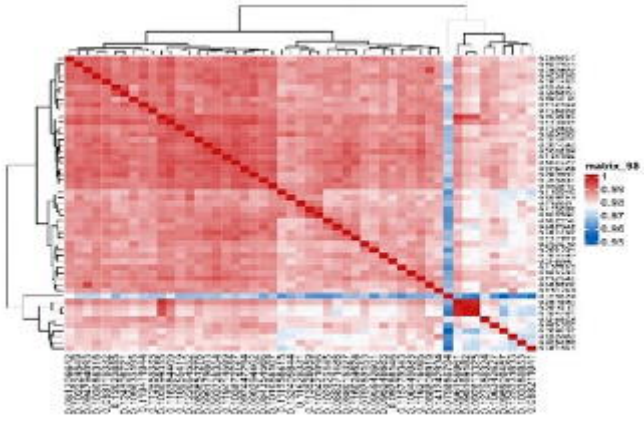


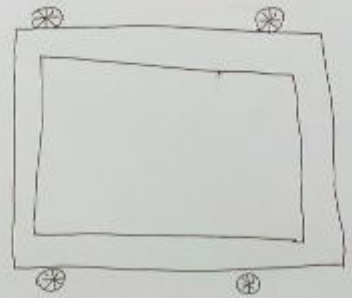


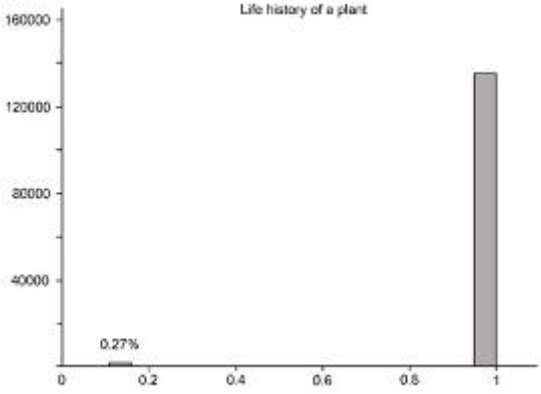


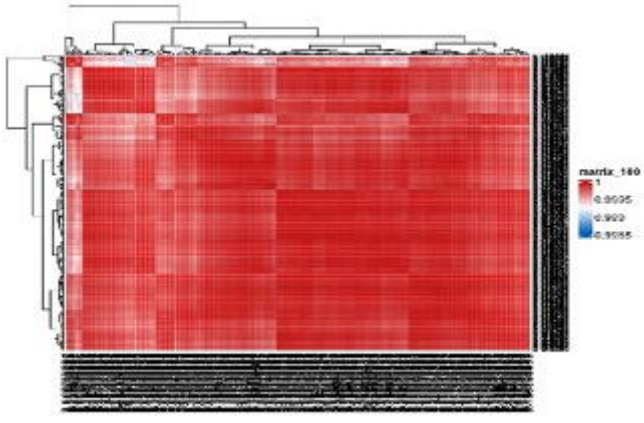
Discovered on
April 5th, 1997


Increasing the carrying capacity of a ship
4000
3000
2000
1000
1.02%
0
0.2
0.4
0.6
0.8
1






Life history of a plant
160000
120000
80000
40000
0.27%
0
0.2
0.4
0.6
0.8
1



Prepare tools

1.准备工具

Measuring cup 量杯

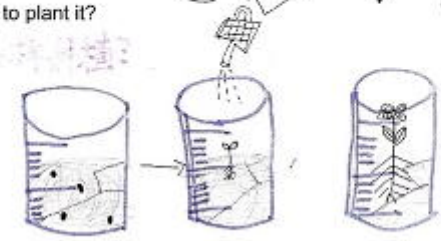
Paper 纸

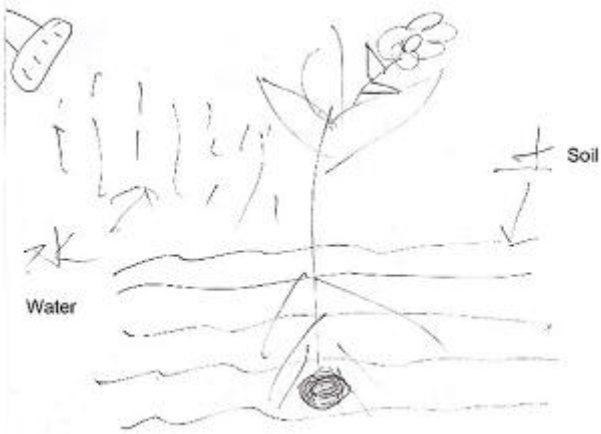
Seed 种子

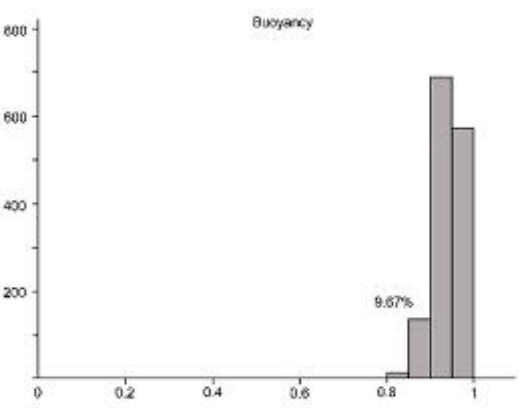
Soil 土

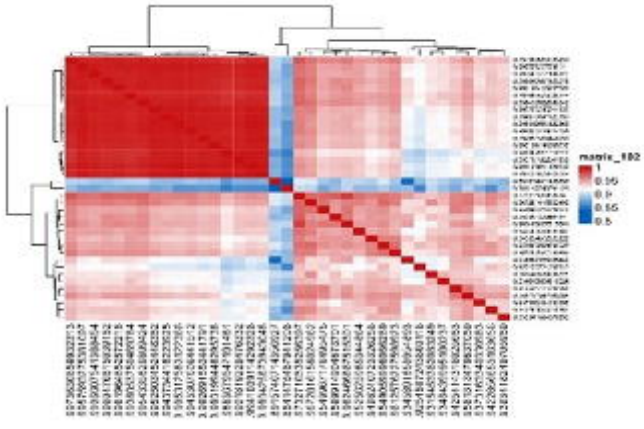
How to plant it?

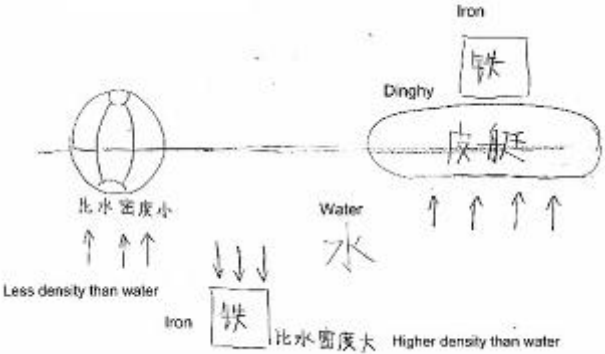
怎样种植?

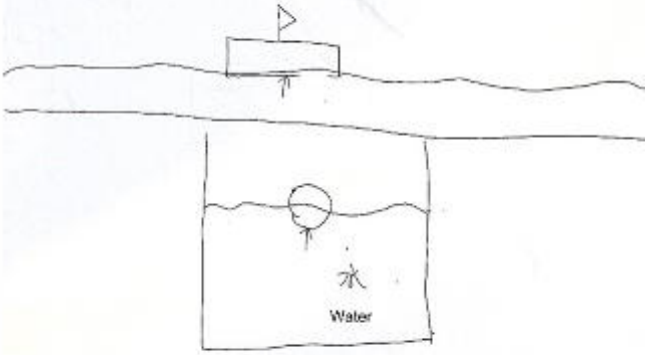


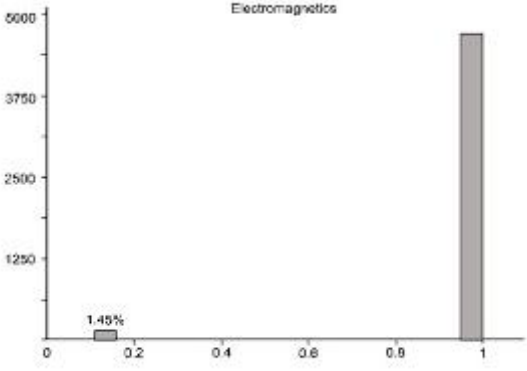


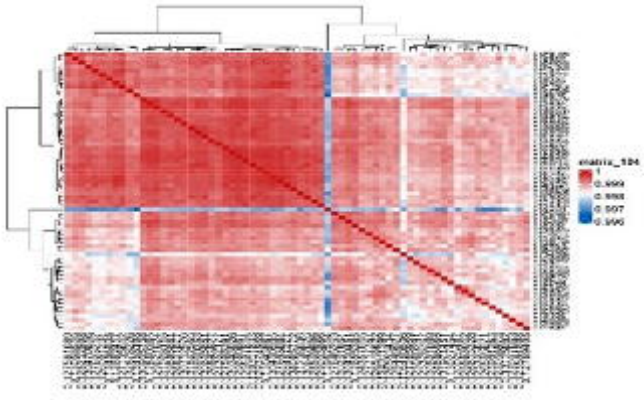

Buoyancy
800
600
400
200
9.67%
0
0.2
0.4
0.6
0.8
1

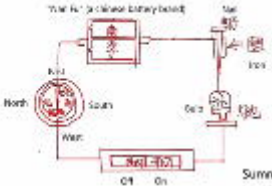

1
0.95
0.9
0.85
0.8


Iron
铁
Dinghy
皮艇
比水密度小
Less density than water
Water
水
Iron
铁
比水密度大
Higher density than water


水
Water


Electromagnetics
5000
3750
2500
1250
1.45%
0
0.2
0.4
0.6
0.8
1



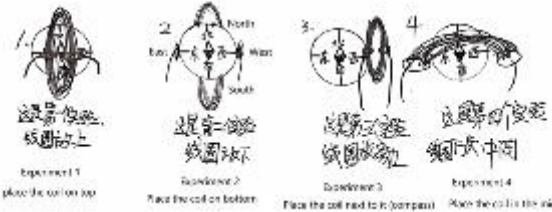


小结：电可以转化为磁，
磁可以转化为电。
这个现方由奥斯特
于1820年偶然发现。
于管能了不起。
自然与磁象可以检
验废电池。

Summary: Electricity can be converted into magnetism, and magnetism can be converted into electricity. This viewpoint was accidentally discovered by Oersted in 1820. Very impressive. This phenomenon can be used to test waste batteries.

这节课，我们学习了"电磁感"，并做了四个实验如下

In this class, we learned about "electromagnetism" and conducted the following four experiments

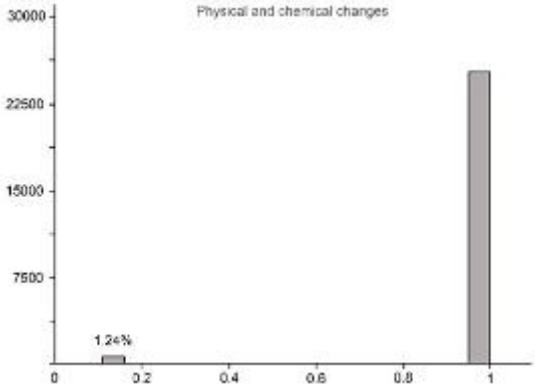


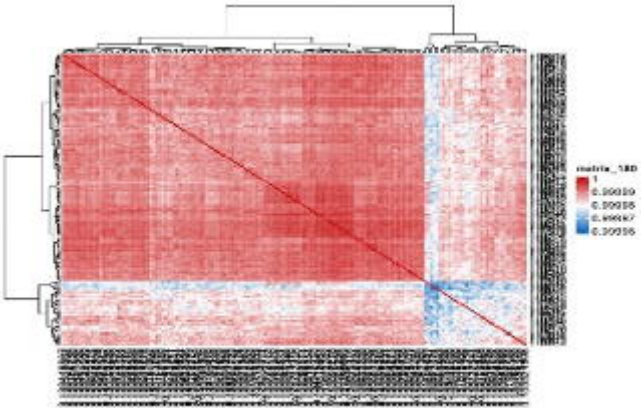
Experiment 1
place the coil on top

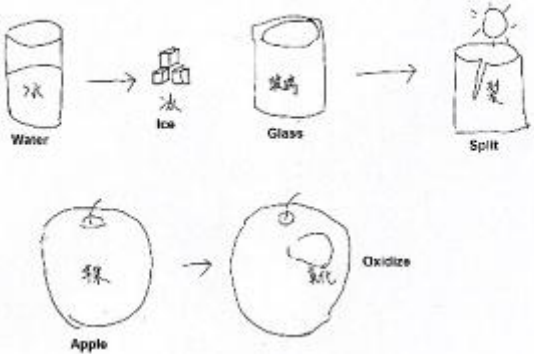
Experiment 2
Place the coil on bottom

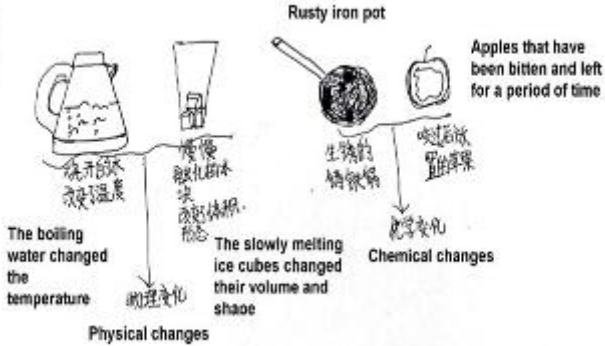
Experiment 3
Place the coil next to it (compass)

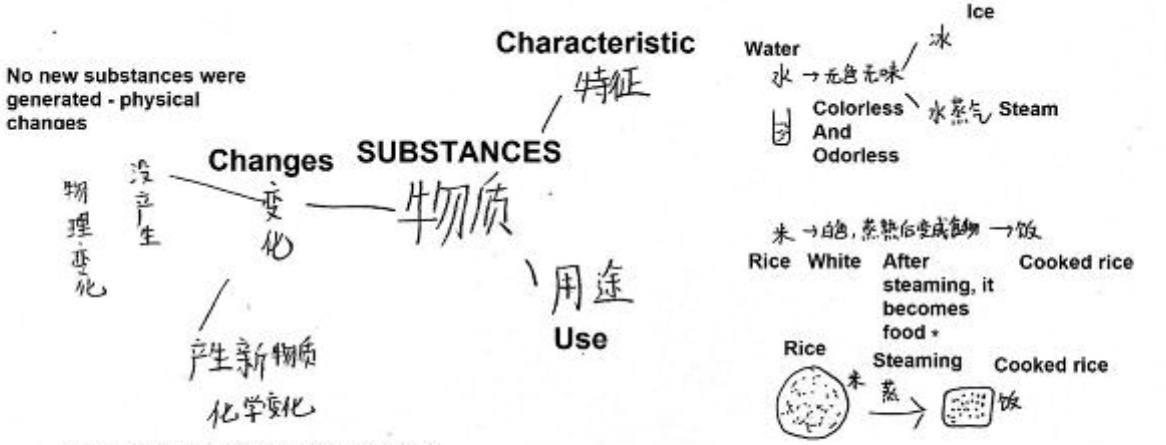
Experiment 4
Place the coil in the middle


Physical and chemical changes
30000
22500
15000
7500
1.24%
0
0.2
0.4
0.6
0.8
1

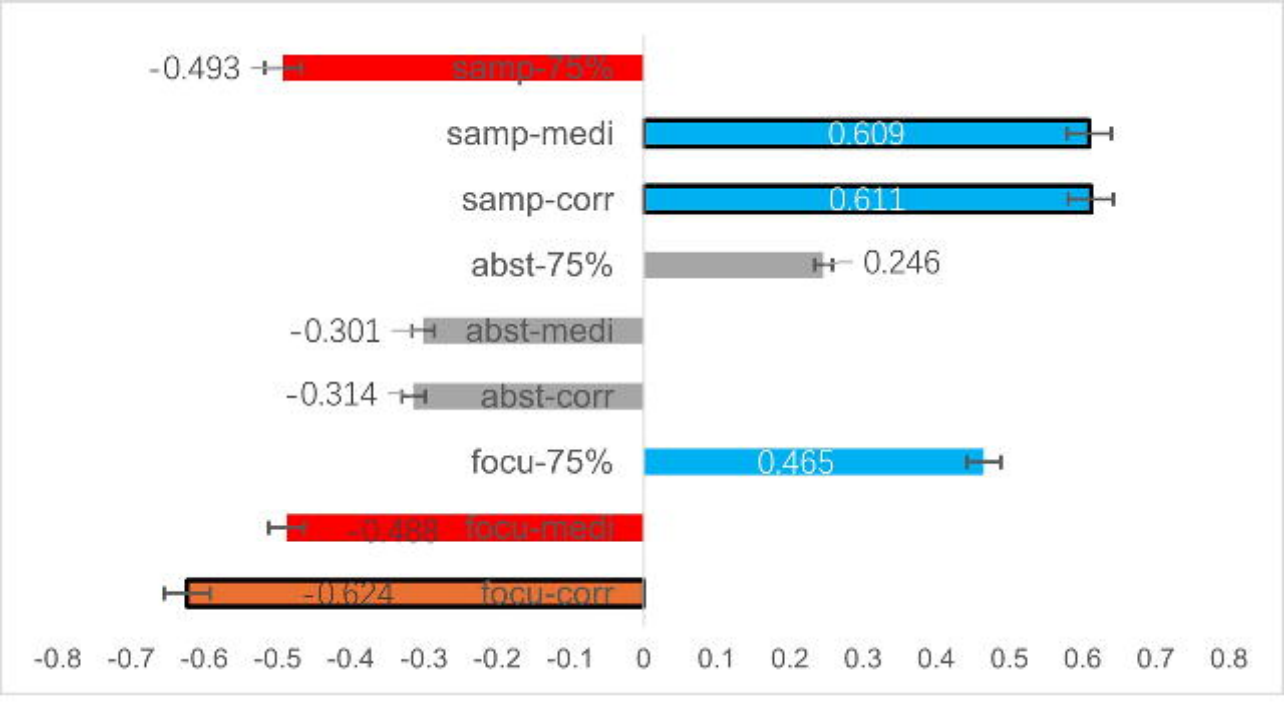



Water
Ice
Glass
Split
Apple
Oxidize


Rusty iron pot
Apples that have been bitten and left for a period of time
The boiling water changed the temperature
The slowly melting ice cubes changed their volume and shape
Chemical changes
Physical changes



* In chinese, raw grain (and rice) and steamed/cooked grain are represented by two completely different words. The all of cooked grain have an unified name, and the name is very similar to that of staple food, so the student here used the word "food".


-0.493
samp-75%
samp-medi
0.609
samp-corr
0.611
abst-75%
0.246
-0.301
abst-medi
-0.314
abst-corr
focu-75%
0.465
-0.488
focu-medi
-0.624
focu-corr
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

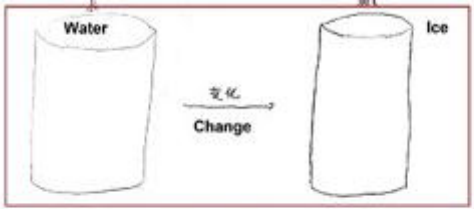
Water will turn into ice when the temperature drops (physical changes)

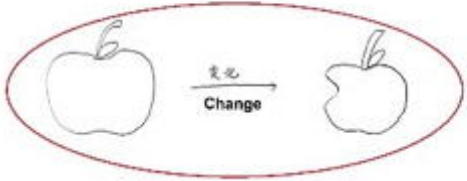
Iron pot rusts (chemical changes)

Oil will solidify into blocks when the temperature decreases* (physical changes)

When apples are bitten, the wound will oxidize (chemical changes)

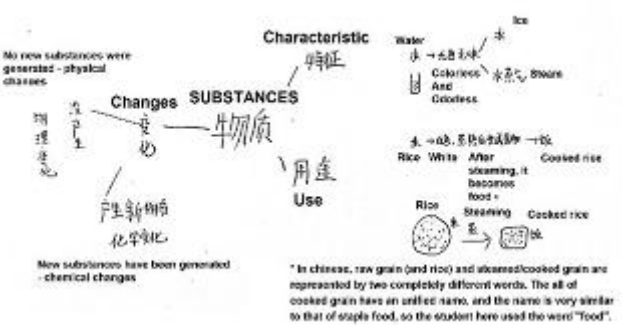
* The original text is" oil will turn into grease when the temperature drops", but the student seems to have not found a suitable vocabulary to describe solidified oil





Boiled water
Melted ice cubes
物理变化
Physical changes

Rusty iron pot
Apple that oxidizes after biting
化学变化
Chemical changes


No new substances were generated - physical changes
Characteristic
特征
Changes
SUBSTANCES
物质
物理变化
产生新物质
化学变化
用途
Use
New substances have been generated - chemical changes
Ice
Water
Colorless And Colorless
Steam
Rice
White
After steaming, it becomes food *
Cooked rice
Steaming
Rice
Cooked rice
* In chinese, raw grain (and rice) and steamed/cooked grain are represented by two completely different words. The all of cooked grain have an unified name, and the name is very similar to that of staple food, so the student here used the word "Food".