Global statistic of students' understandings of Newtonian mechanics represented by a score-state vector of FCI

Masaki Goda [*]
Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2, Nishi-ku, Niigata-City, Niigata 950-2181, Japan

Shinichi Kishizawa [†]
Urawa Akenohoshi Girls' Junior and Senior High School, 9-4-19, Higashiurawa, Midori-ku, Saitama City 336-0926, Japan

Yamato Hasegawa [‡]
Institute of Science Tokyo High School, 3-3-6, Shibaura, Minato-ku, Tokyo 108-0023, Japan

(Received 23 December 2024; accepted 18 June 2025; published 10 July 2025; corrected 13 August 2025)

In this paper, we describe student understanding of the "force concept" (basic concepts of Newtonian mechanics) by representing it as a vector-valued quantity, which we refer to as the "score-state vector," in a (30-dimensional) Force Concept Inventory (FCI) score space. We use a large ensemble of FCI results collected throughout Japan to explore the global statistical structure of the vector. We used principal component analysis to project the score-state vector into a three-dimensional subspace and found that the student score states follow a non-normal distribution. The distribution exhibits two snakes and is also skewed like a screw. This implies that the correlation of the second order (such as the commonly used correlation coefficients) is insufficient to describe the global structure of student score states. We find the distribution to be roughly divided into two subdistributions corresponding to novice respondents and intermediate (transient) and expert respondents. This indicates that "novice" and/or "expert" are not additional qualities given to the ensemble of samples but are understood as a label distinguishing the characteristic areas of the entire distribution. In addition, we observe, as a result of the non-normal distribution, the deviation of the second-order correlation of the score-state vector from that of an artificial interitem-uncorrelated ensemble of samples. The observed deviations realize many previously obtained results and present us with many other new results. Although ours is only a first step, it is a necessary process toward constructing a more accurate model of the student's reasoning about Newtonian mechanics.

DOI: 10.1103/xw7r-2rx1

I. INTRODUCTION

The Force Concept Inventory (FCI) is a widely administered assessment to probe student understanding of the basic concepts of Newtonian mechanics as well as common misconceptions regarding the topic [1,2]. Hestenes *et al.* claimed from their FCI data that getting the concept of force is essential for understanding Newtonian mechanics. Their claim indicates that getting the concept of force is correlated with getting the other basic concepts in Newtonian mechanics [1,2].

Much physics education research has been conducted utilizing pre-post administration of the FCI to compare the effectiveness of various instructional styles. In a seminal report, Hake reported a remarkable difference in student learning between classes utilizing active engagement strategies and traditional lectures [3].

In addition to comparing learning gains, education researchers have used the FCI to study patterns in student responses. These studies have played a role in arguments about how to model the cognitive structure of student ideas. In particular, lack of coherence between responses has been used to argue that many students do not have self-consistent and robust "misconceptions" or "naïve conceptions"; rather their reasoning is better characterized by context-sensitive assemblies of smaller knowledge pieces, referred to as "phenomenological primitives" [4,5], "facets" [6], or "schemas" [7,8].

[*]Contact author: masaki.goda.emeritus@niigata-u.ac.jp
[†]Contact author: s_kishizawa@nifty.com
[‡]Contact author: hasegawa@sth.isct.ac.jp

This distinction is important as it informs instruction [9–18]. For more details, we refer readers to Redish's book [10] and the references within.

An effective discussion regarding students' reasoning, as measured on the FCI, depends upon an accurate interpretation of the FCI results. Usual studies study correlations of responses on the FCI by looking at the second-order correlation coefficients (such as correlation coefficients between two-item scores). This is appropriate when the distribution is normal but insufficient when the distribution is not normal. As we will discuss in this paper, the distribution of FCI responses is not normal, and so we focus our efforts on creating an improved description of the distribution.

One common option for analyzing interitem correlation is factor analysis (FA), a powerful method for discovering hidden factors that affect student response patterns [19–23]. Confirmatory factor analysis (CFA), in particular, is powerful in analyzing the factor and factor relationship [23].

However, the governing equation in FA is linear and contains stochastic variables; hence, it is not suitable for describing the nonlinear dependence of the variable quantities on the student ability (also conceptualized to be the nonlinear response to the student growth) nor to determine the "factor score" (the gains of factors individual student obtains) exactly. Although some methods have been developed to "estimate" the factor score [24], we are unaware of any study that has applied these methods to FCI results.

Item response theory (IRT) describes the nonlinear response by introducing a nonlinear item characteristic function of two distinguished groups of variables: the characteristic variable describing each student (typically proficiency or trait) and those describing each item (such as difficulty and discrimination) [25–27]. The two groups of variables are treated to be independent.

The nominal response model is an expanded version of IRT that describes not only the likelihood of obtaining the correct response on each item but also the likelihood of the respondent selecting each of the other (four) incorrect responses of each item [28,29].

IRT can calculate the proficiency of each student. Multitrait (multivariable or multidimensional) IRT is an extension of IRT that models the likelihood of providing the correct answer to an item when multiple forms of student proficiency are expected to be relevant [30–33].

Since IRT treats students' proficiency and item parameters as independent groups of variables, a given student's proficiency is taken to be constant across items; as such, IRT does not correctly consider the interitem correlation of the individual student's responses.

Hence, despite the great advantage of IRT of taking nonlinear responses reasonably into account, it is not well suited for analyzing correlations between student item responses. Researchers using multitrait IRT have fit the theory to real score data, leading to a variety of interesting results [30–33]; nevertheless, the exact meaning of the obtained correlation is not clear.

In assuming a (for example, logistic) function in modeling data, information about the original data is reasonably taken into account but is not perfect in utilizing IRT.

Morris *et al.* presented an alternative: the item response curve (IRC), which maintains the original information. Specifically, they described the dependence of item success rates on students' abilities by pooling all student responses to each item and creating a corresponding item response curve (IRC) [34,35]. As a proxy for proficiency, they used the simple (arithmetically summed) total score.

While IRC is only a representation of the raw data, it is also a global statistical representation of an ensemble of the students' states.

As we show in this paper, the distribution of the students' score states consisting of 30 item scores of FCI is non-normal (non-Gaussian). The distribution is snaked and skewed. This means that the correlation coefficients usually used are not enough to describe the entire structure of the interitem correlation.

Previous work utilizing IRC has demonstrated global characteristics of the students' responses to the FCI and also the FMCE that are independent of the nationality and learning history [36–39]. As such, we expect that the non-normal distribution we found is not idiosyncratic to Japanese learners but rather a global phenomenon. This is the driving motivation for revisiting the student's state of understanding of Newtonian mechanics. To this end, we return to the representative scheme based upon principal component analysis (PCA) [40–42].

This paper consists of four sections.

In Sec. II, we present data comprised of 13 768 sets of FCI responses (each response comprised in turn of 30 items) collected widely from across Japan. These data are coded as correct or incorrect on each item and assembled into a 30-dimensional vector, which we refer to, in this paper, as a "score-state vector" or "score state" for each student.

We project the score-state vectors into a low-dimensional vector space spanned by the main principal axes of the correlation matrix to reduce the dimensionality without losing the non-normal structural information of the interitem correlation. We find that three dimensions are effective for the first stage of this study.

In the three-dimensional vector space, we observe three item groups corresponding to the five item groups adopted in confirmatory factor analysis (CFA) by Eaton and Willoughby [23].

In Sec. III, we describe the entirety of the distribution of the score-state vectors in the three-dimensional space and observe directly that the entire distribution is not normal.

We try to overcome the difficulty of describing the characteristics of the non-normal distribution in two ways.

The first way is via a primitive but powerful approach of using our eyes to directly identify the characteristic structures of the score-state vectors. In doing so, we find that the distribution snakes twice like a snake and is skewed like a screw.

The second way of describing the higher-order correlations is by observing the deviations of the second-order correlation from that of an ideal inter-item-uncorrelated ensemble of samples. This latter method is also adopted in Sec. II.

The final section, Sec. IV, is devoted to the concluding remarks and discussion.

II. CORRELATION MATRIX AND THE PRINCIPAL AXES

A. Data preparation and the principal component analysis

The data we used for the analysis presented in this paper comprise $N = 13\,768$ FCI respondents from high schools (6720) and universities (7048) throughout Japan. These data were collected in 2014–2016 in a PER project led by Roy Lang and supported by the Grants-in-aid for the Scientific Research 26282032 of the Japan Society for the Promotion of Science (JSPS). The university students were in introductory courses. The high schools and universities span a wide range in terms of prestige and such represent the population of Japanese learners of introductory physics. Most of these data were collected prior to instruction. We intentionally pool these data from diverse learners together to discover the general characteristics of the large sample space.

Prior work investigating the general characteristics of this dataset has found that the IRCs of Japanese high school students and university students are similar; moreover, these IRCs are also similar to those of learners in the United States [36–38].

Provided the total-score distribution extends over the entire range of the score, different ensembles of samples with varying distributions of total score also exhibit similar IRC characteristics. This is because IRC describes a local quantity that does not depend on the intensity of the distribution at the local position. The same is reported of the IRC characteristics of the Force and Motion Conceptual Evaluation (FMCE) [39].

The distribution contains information including both commonalities and individualities between respondents. Different distributions exhibit different characteristics (such as score distribution) depending on the local intensity of the distribution.

However, we find the shared IRC characteristics between populations to be remarkable, and we seek to understand this commonality better. The distribution of the total score of the entire ensemble of samples (named ESJ in this paper) is shown in Fig. 1.

At a glance, we can see that the typical total-score distribution is not normal (not Gaussian).

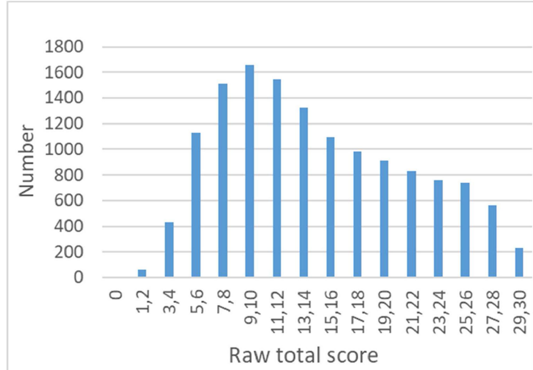


FIG. 1. Number distribution of raw total scores of ESJ of $N = 13\,768$ examinees.

The average rate of answering correctly and the standard deviation of each item's score (taking the value of 1 (correct) or 0 (incorrect) for each student) are shown in Table I. The average total-score rate of ESJ is about 14/30.

Although we want to investigate the student response to items, we start by investigating the correct (or incorrect) answer in this paper.

We define here a 30-dimensional (row) vector $\mathbf{s}_n^0$ named as (entire) score-state vector of examinee $n(n = 1 - N)$ as,

$$\mathbf{s}_n^0 = \left(s_{1,n}^0,\, s_{2,n}^0,\, s_{3,n}^0,, \ldots,\, s_{30,n}^0\right), \tag{1}$$

where the kth component (code) of which (denoted as $s_{k,n}^0$) takes the value of the item score of 1 (correct) or 0 (incorrect) depending on the nth examinee. Obviously, the kth component, $s_{k,n}^0$, of $\mathbf{s}_n^0$ is the kth item score of the examinee n.

TABLE I. Score rate (average rate of answering correctly) and the standard deviation σ of 30 items (questions).

Item no.	Score rate	σ	Item no.	Score rate	σ
1	0.6292	0.4830	16	0.5308	0.4991
2	0.5285	0.4992	17	0.2788	0.4484
3	0.7471	0.4347	18	0.3880	0.4873
4	0.3109	0.4629	19	0.5763	0.4942
5	0.3032	0.4596	20	0.5063	0.5000
6	0.6843	0.4648	21	0.3784	0.4850
7	0.5556	0.4969	22	0.4747	0.4994
8	0.5960	0.4907	23	0.3872	0.4871
9	0.4869	0.4998	24	0.6291	0.4831
10	0.7154	0.4513	25	0.3552	0.4786
11	0.3600	0.4800	26	0.2288	0.4201
12	0.7194	0.4493	27	0.5612	0.4963
13	0.3347	0.4719	28	0.5559	0.4969
14	0.3954	0.4889	29	0.6469	0.4780
15	0.3067	0.4611	30	0.2948	0.4560

It is important to note that the score-state vector $\mathbf{s}_n^0$ is a vector-valued quantity having full information of the 30-item scores of examinee n.

The score-state vector $\mathbf{s}_n^0$ is then normalized to make the normalized score-state vector $\mathbf{s}_n$.

That is, each kth component of $\mathbf{s}_n^0$ (that is $s_{k,n}^0$) is normalized by the kth component of the mean vector $\langle \mathbf{s}_n^0 \rangle$ and by the kth component of the corresponding standard deviation vector $\boldsymbol{\sigma}^0$ (that is σ_k^0) to make the normalized score-state vector $\mathbf{s}_n$ of examinee n.

This normalization is made to discard nonessential factors that may be induced in the process of item (question) making. We call the kth component of $\mathbf{s}_n$ (that is $s_{k,n}$) normalized kth item score of examinee n.

Note that the normalized score-state vector $\mathbf{s}_n$ has full information of the correct-incorrect answers of 30 items of the examinee n.

We can say $\mathbf{s}_n$ represents the score state or, more simply, the "state" of the nth examinee.

Of course, 30 dimensions are too many to discuss the structure of the examinees' states. This is the reason why we utilize PCA. We want to reduce the dimensions of the state of the examinee without losing the essential part of the entire structure of the correlation in the distribution. This point is considered later carefully in this section.

By accumulating the normalized score-state vectors (hereafter, we omit the term "normalized" throughout this paper and refer them only as score-state vectors, etc.) of N examinees, we make an N-row $\times$30-column data matrix $\mathbf{D}$ [40], where the n, k element $s_{k,n}$ is the kth code of the score-state vector $\mathbf{s}_n$.

As score-state vectors $\mathbf{s}_n (n = 1 - N)$ have full information on the correct-incorrect answers of FCI, the same is true for the data matrix $\mathbf{D}$.

The distribution of the N students' score states in the 30-dimensional score space is identical to the data matrix $\mathbf{D}$. Thus, the distribution of the N students' score states is also the most fundamental quantity having full information of the students' correct-incorrect responses to FCI.

By using the data matrix, we first numerically calculate a real and symmetric correlation matrix $\mathbf{C}$ of 30×30 dimension, where the k, k' element is the (Pearson's) correlation coefficient of correct answers between k and k' items.

$$c_{k,k'} = \frac{\sum_{n=1}^{N} (s_{k,n}^0 - \langle s_{k,n}^0 \rangle)(s_{k',n}^0 - \langle s_{k',n}^0 \rangle)}{N \sigma_k \sigma_{k'}} = \frac{1}{N} \sum_{n=1}^{N} (s_{k,n} \cdot s_{k',n}), \qquad (k, k' = 1 - 30). \tag{2}$$

Note that the correlation coefficient defined in Eq. (2) is the correlation of the second order in the sense that it is the correlation between two quantities (of two items).

As one of the purposes of this paper is to see the deviation of the observable quantities from those of inter-item-uncorrelated reference systems, we also make, as reference systems, two artificial ensembles of N random samples, named R0 and R1, in the 30-dimensional item score space.

All of the components of the data matrix $\mathbf{D}_{R0}$ are random values that take the value of 1 or 0 with a "mean" mv fixed to the value of about $14/30$ independent of the items, for all of the N artificial examinees. A large number of the set of artificially made N students' samples constitutes the ensemble of the numerical distributions R0.

On the other hand, all of the components of the data matrix $\mathbf{D}_{R1}$ are random values taking the value of 1 or 0 with a "mean" mv independent of items, but the value of mv distributes depending on each artificial examinee to make the mv distribution corresponding to the total-score distribution as that of ESJ shown in Fig. 1.

In this case, the artificially generated large number of distributions, R1, has the same (artificial) total-score distribution as that of ESJ, though there is no interitem correlation for individual artificial students.

The ensemble R0 has been used for finding the number of hidden factors in real systems in a process referred to as parallel analysis [43].

However, we introduce the ensemble R1 in this paper to compare the set of our real samples ESJ with the ideal (for individuals) inter-item-uncorrelated set of samples having the same total-score distribution as that of ESJ. This is a new version of the parallel analysis.

While correlation coefficients between items or between factors have been discussed in many ways [19–23], we think it is important to observe the deviation of the correlation from that of the inter-item-uncorrelated (for individuals) correlation of R1 having the same total-score distribution as that of ESJ.

Note that R1 differs from R0 in that the interitem correlation coefficients (i.e., the off-diagonal elements of the correlation matrix $\mathbf{C}_{R1}$) do not vanish. This is due to the non-Gaussian total-score distribution shown in Fig. 1. The broad total-score distribution causes uniform positive interitem correlation coefficients even in the (for individuals) inter-item-uncorrelated case.

A numerical sample of $\mathbf{C}_{R1}$ is shown in Fig. 1 in Supplemental Material with the characteristics listed in Table I in Supplemental Material [44].

The off-diagonal elements of the correlation matrix $\mathbf{C}_{R0}$ of R0 are distributed around $0.000 \pm \sigma_{R0}$ and those of $\mathbf{C}_{R1}$ of R1 are distributed around $0.183 \pm \sigma_{R1}$, where $\sigma_{R0} \simeq \sigma_{R1} = 0.008$ are the standard deviations of the randomly fluctuated off-diagonal elements (interitem correlation coefficients) of $\mathbf{C}_{R0}$ and $\mathbf{C}_{R1}$, respectively.

The correlation matrix $\mathbf{C}$ (corresponding to the real data) is shown in Fig. 2 with the item numbers arranged in a way to ease reference for discussion below.

		EW12 Item Group														EW3 Item Group				EW45 Item Group								EWe			
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26	Q1	Q2	Q3	Q29
	Q6	1	0.24	0.15	0.21	0.16	0.19	0.18	0.13	0.17	0.16	0.14	0.11	0.16	0.14	0.13	0.09	0.14	0.16	0.11	0.15	0.2	0.14	0.18	0.12	0.15	0.15	0.15	0.13	0.1	0.03
	Q7	0.24	1	0.22	0.2	0.2	0.2	0.16	0.21	0.18	0.23	0.18	0.14	0.18	0.15	0.16	0.12	0.21	0.18	0.19	0.19	0.26	0.21	0.22	0.17	0.2	0.21	0.19	0.13	0.12	0.06
	Q8	0.15	0.22	1	0.29	0.26	0.3	0.22	0.4	0.23	0.28	0.22	0.14	0.21	0.22	0.18	0.14	0.25	0.23	0.22	0.24	0.29	0.25	0.25	0.2	0.23	0.24	0.2	0.14	0.17	0.08
	Q10	0.21	0.2	0.29	1	0.29	0.24	0.33	0.23	0.26	0.26	0.27	0.14	0.22	0.23	0.2	0.17	0.29	0.3	0.19	0.28	0.29	0.21	0.24	0.18	0.22	0.23	0.27	0.17	0.2	0.1
	Q20	0.16	0.2	0.26	0.29	1	0.24	0.25	0.27	0.19	0.27	0.38	0.2	0.25	0.22	0.19	0.12	0.24	0.26	0.2	0.25	0.3	0.22	0.27	0.19	0.24	0.26	0.24	0.15	0.19	0.07
	Q23	0.19	0.2	0.3	0.24	0.24	1	0.41	0.25	0.21	0.3	0.22	0.26	0.32	0.23	0.16	0.1	0.21	0.21	0.22	0.22	0.3	0.22	0.31	0.2	0.24	0.28	0.19	0.11	0.16	0.02
EW12	Q24	0.18	0.16	0.22	0.33	0.25	0.41	1	0.21	0.21	0.23	0.24	0.17	0.25	0.22	0.16	0.12	0.22	0.27	0.17	0.22	0.26	0.18	0.24	0.17	0.22	0.24	0.23	0.13	0.17	0.06
	Q9	0.13	0.21	0.4	0.23	0.27	0.25	0.21	1	0.19	0.3	0.26	0.2	0.22	0.2	0.18	0.1	0.24	0.22	0.23	0.24	0.3	0.24	0.26	0.21	0.25	0.27	0.19	0.11	0.15	0.04
	Q12	0.17	0.18	0.23	0.26	0.19	0.21	0.21	0.19	1	0.22	0.19	0.14	0.18	0.19	0.15	0.12	0.2	0.2	0.16	0.19	0.25	0.18	0.19	0.15	0.18	0.18	0.18	0.13	0.13	0.07
	Q14	0.16	0.23	0.28	0.26	0.27	0.3	0.23	0.3	0.22	1	0.26	0.24	0.26	0.24	0.2	0.13	0.26	0.23	0.25	0.25	0.33	0.25	0.29	0.23	0.29	0.29	0.23	0.15	0.17	0.03
	Q19	0.14	0.18	0.22	0.27	0.38	0.22	0.24	0.26	0.19	0.26	1	0.19	0.23	0.22	0.17	0.1	0.21	0.24	0.18	0.22	0.26	0.18	0.23	0.14	0.21	0.23	0.23	0.13	0.18	0.06
	Q21	0.11	0.14	0.14	0.14	0.2	0.26	0.17	0.2	0.14	0.24	0.19	1	0.26	0.18	0.11	0.05	0.15	0.16	0.16	0.15	0.2	0.14	0.2	0.13	0.17	0.21	0.13	0.08	0.12	0.03
	Q22	0.16	0.18	0.21	0.22	0.25	0.32	0.25	0.22	0.18	0.26	0.23	0.26	1	0.21	0.17	0.1	0.23	0.2	0.21	0.22	0.29	0.21	0.28	0.2	0.25	0.33	0.19	0.12	0.18	0.02
	Q27	0.14	0.15	0.22	0.23	0.22	0.23	0.22	0.2	0.19	0.24	0.22	0.18	0.21	1	0.13	0.06	0.18	0.2	0.16	0.18	0.23	0.16	0.22	0.13	0.19	0.22	0.16	0.09	0.16	0.06
	Q4	0.13	0.16	0.18	0.2	0.19	0.16	0.16	0.18	0.15	0.2	0.17	0.11	0.17	0.13	1	0.29	0.26	0.35	0.23	0.2	0.29	0.22	0.25	0.22	0.23	0.24	0.2	0.16	0.13	0.05
EW3	Q15	0.09	0.12	0.14	0.17	0.12	0.1	0.12	0.1	0.12	0.13	0.1	0.05	0.1	0.06	0.29	1	0.25	0.26	0.17	0.14	0.22	0.17	0.17	0.18	0.16	0.14	0.14	0.12	0.08	0.05
	Q16	0.14	0.21	0.25	0.29	0.24	0.21	0.22	0.24	0.2	0.26	0.21	0.15	0.23	0.18	0.26	0.25	1	0.3	0.26	0.29	0.34	0.3	0.3	0.34	0.36	0.3	0.23	0.16	0.16	0.1
	Q28	0.16	0.18	0.23	0.3	0.26	0.21	0.27	0.22	0.2	0.23	0.24	0.16	0.2	0.2	0.35	0.26	0.3	1	0.22	0.23	0.3	0.23	0.26	0.22	0.26	0.27	0.23	0.16	0.17	0.11
	Q5	0.11	0.19	0.22	0.19	0.2	0.22	0.17	0.23	0.16	0.25	0.18	0.16	0.21	0.16	0.23	0.17	0.26	0.22	1	0.33	0.45	0.48	0.39	0.3	0.33	0.32	0.17	0.17	0.12	0.04
	Q11	0.15	0.19	0.24	0.28	0.25	0.22	0.22	0.24	0.19	0.25	0.22	0.15	0.22	0.18	0.2	0.14	0.29	0.23	0.33	1	0.41	0.35	0.37	0.26	0.31	0.3	0.19	0.15	0.17	0.08
	Q13	0.2	0.26	0.29	0.29	0.3	0.3	0.26	0.3	0.25	0.33	0.26	0.2	0.29	0.23	0.29	0.22	0.34	0.3	0.45	0.41	1	0.46	0.55	0.36	0.41	0.42	0.26	0.23	0.19	0.05
EW45	Q18	0.14	0.21	0.25	0.21	0.22	0.22	0.18	0.24	0.18	0.25	0.18	0.14	0.21	0.16	0.22	0.17	0.3	0.23	0.48	0.35	0.46	1	0.4	0.3	0.32	0.32	0.19	0.17	0.12	0.08
	Q30	0.18	0.22	0.25	0.24	0.27	0.31	0.24	0.26	0.19	0.29	0.23	0.2	0.28	0.22	0.25	0.17	0.3	0.26	0.39	0.37	0.55	0.4	1	0.32	0.38	0.38	0.22	0.18	0.14	0.01
	Q17	0.12	0.17	0.2	0.18	0.19	0.2	0.17	0.21	0.15	0.23	0.14	0.13	0.2	0.13	0.22	0.18	0.34	0.22	0.3	0.26	0.36	0.3	0.32	1	0.44	0.36	0.18	0.15	0.09	0.04
	Q25	0.15	0.2	0.23	0.22	0.24	0.24	0.22	0.25	0.18	0.29	0.21	0.17	0.25	0.19	0.23	0.16	0.36	0.26	0.33	0.31	0.41	0.32	0.38	0.44	1	0.5	0.21	0.16	0.13	0.04
	Q26	0.15	0.21	0.24	0.23	0.26	0.28	0.24	0.27	0.18	0.29	0.23	0.21	0.33	0.22	0.24	0.14	0.3	0.27	0.32	0.3	0.42	0.32	0.38	0.36	0.5	1	0.23	0.19	0.16	0.03
	Q1	0.15	0.19	0.2	0.27	0.24	0.19	0.23	0.19	0.18	0.23	0.23	0.13	0.19	0.16	0.2	0.14	0.23	0.23	0.17	0.19	0.26	0.19	0.22	0.18	0.21	0.23	1	0.3	0.16	0.05
EWe	Q2	0.13	0.13	0.14	0.17	0.15	0.11	0.13	0.11	0.13	0.15	0.13	0.08	0.12	0.09	0.16	0.12	0.16	0.16	0.17	0.15	0.23	0.17	0.18	0.15	0.16	0.19	0.3	1	0.11	0.03
	Q3	0.1	0.12	0.17	0.2	0.19	0.16	0.17	0.15	0.13	0.17	0.18	0.12	0.18	0.16	0.13	0.08	0.16	0.17	0.12	0.17	0.19	0.12	0.14	0.09	0.13	0.16	0.16	0.11	1	0.06
	Q29	0.03	0.06	0.08	0.1	0.07	0.02	0.06	0.04	0.07	0.03	0.06	0.03	0.02	0.06	0.05	0.05	0.1	0.11	0.04	0.08	0.05	0.08	0.01	0.04	0.04	0.03	0.05	0.03	0.06	1

FIG. 2. Correlation matrix C. The order of item (Q) number has been arranged. The warm colors are put on the correlation coefficients that deviate from the mean value 0.183 of the inter-item-uncorrelated (for individuals) correlation matrix $\mathbf{C}_{R1}$ positively by more than double the standard deviation 0.008 of $\mathbf{C}_{R1}$. The cool colors with lateral stripes are put on them, which deviate from the mean of 0.183 negatively by more than double of 0.008.

In Fig. 2, noteworthy deviations of the correlation coefficients from the mean value of those of the inter-item-uncorrelated set of samples R1 are colored.

Warm colors represent positive deviations by more than double the standard deviation of 0.008 from the mean of $\mathbf{C}_{R1}$ (0.183). Cool colors with lateral stripes represent negative deviations by more than double the standard deviation from the mean of $\mathbf{C}_{R1}$.

Thus, the correlation coefficients of $\mathbf{C}$ with warm colors represent the statistical coherence between responses of items for individual students. Those with cool colors with lateral stripes represent statistical confusion or conflict between the responses of items for individual students.

Note that there is a negative deviation of the interitem correlations between Q15 and the items in the EW12 item group, consistent with prior literature [45]. Item Q29 has a negative deviation against all other items, though we do not mention more about it in this paper.

We next numerically diagonalize the correlation matrix $\mathbf{C}$ and get the eigenvalues λ_i $(i = 1 - 30)$ and the eigenvectors (column vectors) $\mathbf{v}_i$ $(i = 1 - 30)$.

$$\mathbf{C}\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i, \qquad (i = 1 - 30). \tag{3}$$

The eigenvalues are shown in Fig. 3 with the reference eigenvalues of $\mathbf{C}_{R1}$.

The eigenvectors $\mathbf{v}_i (i = 1 - 30)$ are the most fundamental (orthogonal and normalized) set of bases describing the

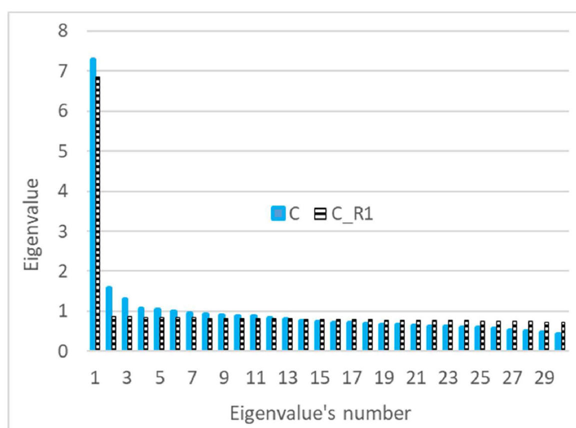


FIG. 3. Eigenvalues of the correlation matrix $\mathbf{C}$ in decreasing order, with the referential values of the inter-item-uncorrelated correlation matrix $\mathbf{C}_{R1}$.

30-dimensional vector space. In what follows, we refer to them as "principal axes."

The ith principal-axis component of any vector is the projection of the vector to the ith principal axis, and we call it the ith principal component (i PC) of the vector. The principal component (PC) is useful to represent the score-state vector $\mathbf{s}_n$.

Projection is only a process of representing a vector and does not modify or deform the original information of the vector. When we project the score-state vector to the 30 principal axes, nothing has changed at all. Even when we project it to some restricted number of principal axes, that is, to the subspace spanned by the corresponding principal axes, it does not modify or deform the information contained within the vector, though some (mainly random) information disappears.

A key concept of PCA is that the variance of the distribution along a principal axis measures the information of the distribution along the axis.

As the variances of the distribution are the eigenvalues themselves corresponding to the principal axes, we can understand that when we have the eigenvalues, such as those seen in Fig. 3, the main systematic information must exist in the subspace spanned by some principal axes having distinguished eigenvalues.

The three main eigenvectors of $\mathbf{C}$ with the corresponding distinguished eigenvalues are listed in decreasing order of the eigenvalues in Table II.

The eigenvalues of $\mathbf{C}_{R1}$ in Fig. 3 have a striking feature that only a single eigenvalue is prominent, reflecting the fact that the total score is the only characteristic value. All other eigenvalues less than 1 describe the information from random fluctuation of the item scores.

The eigenvalue λ_i of $\mathbf{C}$ which is bigger than that of $\mathbf{C}_{R1}$ (see Fig. 3) has a positive amount of information on the "interitem correlation" along the ith principal axis.

With these points in mind, we see that the distribution of ESJ along the first principal axis corresponds to the total score, while the first three principal axes combined contain information on the interitem correlation.

Although additional information is contained within the remaining axes, as a first step, we restrict our analysis in this paper to the three main principal axes.

The degree of freedom for describing the factor structure has meaning in FA. The factor spaces of different dimensions describe different worlds. Five or six factors are frequently discussed [20–23], but a bifactor structure is also suggested [32].

In PCA, however, the number of principal axes is not essential. Even when we consider axes beyond the first three, all of the results obtained in analyzing the three-dimensional space are unaffected.

TABLE II. Three main eigenvectors (Eig.s) of the correlation matrix $\mathbf{C}$. The corresponding eigenvalues are, $\lambda_1 = 7.282$, $\lambda_2 = 1.578$, and $\lambda_3 = 1.297$.

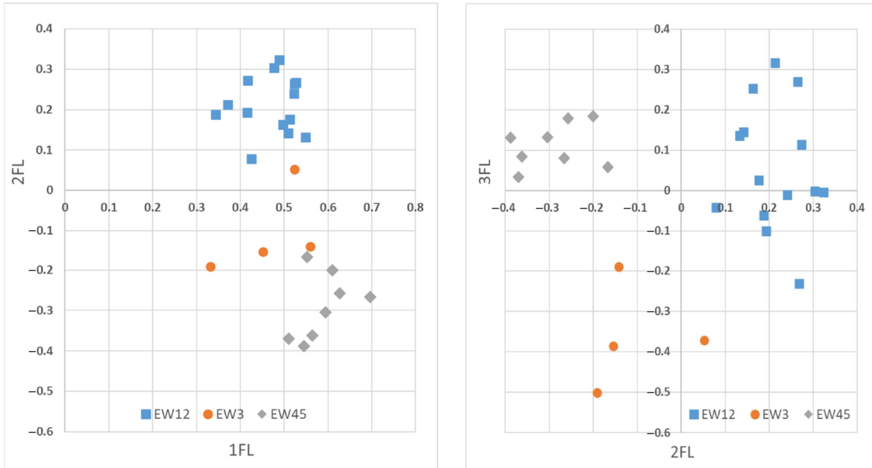
	Eig. 1	Eig. 2	Eig. 3
1	0.1681	0.1091	−0.2249
2	0.1255	−0.0363	−0.2364
3	0.1242	0.1926	−0.0704
4	0.1672	−0.1234	−0.3385
5	0.2019	−0.3091	0.1151
6	0.1274	0.1490	−0.0541
7	0.1574	0.0628	−0.0366
8	0.1902	0.1406	0.0227
9	0.1889	0.1130	0.1275
10	0.1953	0.2131	−0.2024
11	0.2048	−0.1332	0.0511
12	0.1539	0.1538	−0.0878
13	0.2579	−0.2122	0.0709
14	0.2035	0.1053	0.1201
15	0.1229	−0.1526	−0.4400
16	0.2074	−0.1130	−0.1667
17	0.1893	−0.2949	0.0297
18	0.2093	−0.2887	0.0746
19	0.1768	0.2419	−0.0013
20	0.1938	0.1914	−0.0096
21	0.1376	0.1691	0.2782
22	0.1843	0.1299	0.2228
23	0.1939	0.2102	0.2372
24	0.1812	0.2577	−0.0034
25	0.2204	−0.2425	0.1162
26	0.2261	−0.1593	0.1620
27	0.1547	0.2174	0.1003
28	0.1941	0.0413	−0.3258
29	0.0456	0.0673	−0.2774
30	0.2324	−0.2049	0.1575

In the following Secs. II B and II C, we provide additional justification from our data for our choice to focus on the first three primary axes.

B. Three item groups

To confirm the validity of the three-dimensional PC space and also to understand the interitem correlation more visually, we start by considering the ith factor-loading (FL) vector $\mathbf{f}_i = \sqrt{\lambda_i}\mathbf{v}_i$ ($i = 1 - 30$). The kth code (component) of that vector describes the correlation coefficient between the kth component of score-state vectors $\mathbf{s}_n$ ($n = 1 - N$) (that is, the kth item scores of N students) and the ith PCs of score-state vectors $\mathbf{s}_n$ ($n = 1 - N$) [46]. See Eq. (5) in the next subsection.

The FL-FL plots between $\mathbf{f}_1$, $\mathbf{f}_2$, and $\mathbf{f}_3$ are shown in Fig. 4.



(a) (b)

FIG. 4. Factor-loading (FL)—factor-loading (FL) plots of the 26 items. (a) Second factor-loading (2 FL) vs first factor-loading (1 FL) plot. (b) Third factor-loading (3 FL) vs second factor-loading (2 FL) plot. The blue squares, orange circles, and silver diamonds represent, respectively, the items of EW12, EW3, and EW45 item groups.

The FL-FL plots describe an interrelation between items.

$$\begin{aligned}&\sum_{i=1}^{30}(f_{k,i}-f_{k',i})^2\\&=\sum_{i=1}^{30}(f_{k,i})^2+\sum_{i=1}^{30}(f_{k',i})^2-2\sum_{i=1}^{30}(f_{k,i}f_{k',i})\\&=2(1-c_{k,k'}),\end{aligned} \tag{4}$$

where, $f_{k,i}$ is the kth cord of the ith FL vector $\mathbf{f}_i$.

The first term of the second equation is the correlation coefficient $c_{k,k}$, the second term of it is $c_{k',k'}$, and the third term of it is (−2) times of $c_{k,k'}$. The relation between FL vectors and the correlation coefficients is in the spectral decomposition of the correlation matrix $\mathbf{C}$ given in Eq. (5) in the next subsection.

Equation (4) indicates that when the correlation between any item pair is strong, the distance between them is close in the FL space. From Eq. (4) and the large amount of systematic information on the interitem correlation in the three-dimensional PC space, we use the projection of the items to the three-dimensional FL space to interpret the relation between items.

In Fig. 4, we plot three item groups corresponding to those adopted by Eaton and Willoughby in their confirmatory factor analysis (CFA) [23]. Eaton and Willoughby adopted five item groups (named $\mathrm{EW1}-\mathrm{EW5}$ in this paper), with the exceptional four items (named EWe in this paper), in their CFA analysis. The item list of their five item groups is shown in Table III.

In our FL-FL plot of ESJ in Fig. 4, we plot three item groups $\mathrm{EW12}\equiv\mathrm{EW1}+\mathrm{EW2}$, EW3, and $\mathrm{EW45}\equiv\mathrm{EW4}+\mathrm{EW5}$.

When we extend the dimensions of the PC space to 5 by adding the 4th and 5th principal axes (using $\mathbf{f}_4$ and $\mathbf{f}_5$), EW45 splits into EW4 and EW5, but we could not confirm the split of EW12 into EW1 and EW2.

This fact tells us that the results of PCA are very close to those of CFA in the case of FCI, though not identical. The results are particularly aligned when the stochastic character in their structural equation model is insignificant in FA.

EW12 item group consists of seven items on Newton's first law and seven items on Newton's second law. The 14 items have a good proportion in the 30 FCI items, corresponding to the fact that the body of the introductory physics class may be the study of Newton's first and second laws, including the equation of motion.

EW3 item group consists of four items on Newton's third law (action-reaction law).

EW45 item group consists of eight items asking, in various kinds of situations, what kind of force is acting on.

The force is formally defined in the second law, but Hestenes *et al.* discerned the important position of understanding the concept of force in Newtonian mechanics and made the eight items concentrated on asking about the concept of force in various situations.

TABLE III. Five item groups and the exception adopted by Eaton and Willoughby [23].

Group	Item no.	Characterization
EW1	6, 7, 8, 10, 20, 23, 24	First law with kinematics
EW2	9, 12, 14, 19, 21, 22, 27	Second law with kinematics
EW3	4, 15, 16, 28	Third law
EW4	5, 11, 13, 18, 30	Force identification
EW5	17, 25, 26	Force identification (mixed)
EWe	1, 2, 3, 29	(Exception)

We have shown the deviation of the correlation matrix $\mathbf{C}$ from $\mathbf{C}_{R1}$ in Fig. 2 by changing the order of the item numbers to one suitable for representing these item groups.

We find remarkable intra- and intercorrelation in and between the item groups EW12, EW3, and EW45, upon which we now elaborate.

IIB.1. The EW45 item group shows a comparatively high positive deviation in the intragroup correlation, exhibiting good coherence in the individual's understanding of the eight items of EW45 group on the concept of force.

IIB.2. There is also a positive deviation of the correlation between EW45 item group and both of EW12 and EW3 groups, suggesting a coherence between the individual's understanding of the items of EW45 on the concept of force and that of the other items of EW12 and EW3 covering the first, second, and third laws. Considering IIB.1 and Table II, the statement IIB.2 is consistent with the claims of Hestenes *et al.* that the concept of force is central to Newtonian mechanics [1,2].

IIB.3. The deviation of the correlation between the EW3 group (Newton's third law) and the EW12 group (Newton's first and second laws) is more nuanced.

Specifically, items 15 and 4, in EW3, show a negative deviation of the correlation with the items of EW12, while items 16 and 28 exhibit a positive deviation.

The positive deviation with item 16 represents the conflation that many students make between Newton's third law and Newton's second law when forces are balanced.

These confusions have been discussed extensively by Law and Wilson [45]. We refer to these confusions as the N2-NF (net force of Newton'ssecond law) and N3 (Newton's third law) confusions and will return to them again in statement IIC.b2 of the next subsection and in statement IIIC.b3 of the next section.

C. PC expansion of the correlation matrix

To explore the internal structure of the correlation matrix $\mathbf{C}$ in the three-dimensional PC space and to confirm the validity of the three-dimensional PC space, we make a PC expansion of the correlation matrix $\mathbf{C}$. The correlation matrix $\mathbf{C}$ is spectrally decomposed into 30 terms as a function of the eigenvalues and eigenvectors as [41],

$$\begin{aligned}\mathbf{C} &= \sum_{i=1}^{30} \mathbf{v}_i \lambda_i (\mathbf{v}_i)^t \\ &= \mathbf{v}_1 \lambda_1 (\mathbf{v}_1)^t + \mathbf{v}_2 \lambda_2 (\mathbf{v}_2)^t + \mathbf{v}_3 \lambda_3 (\mathbf{v}_3)^t + \cdots \\ &= \sum_{i=1}^{30} \mathbf{f}_i (\mathbf{f}_i)^t \Rightarrow \sum_{i=1}^{30} \mathbf{C}_i, \end{aligned} \tag{5}$$

where $(\mathbf{v}_i)^t$ is the transposed (row) vector of $\mathbf{v}_i$.

The first term $\mathbf{C}_1$ represents the first contribution to the correlation coefficients from the distribution of $\mathbf{a}_n (n = 1 - N)$ along the first principal axis. The second term $\mathbf{C}_2$ represents the second contribution from the distribution along the second principal axis, and the third term $\mathbf{C}_3$ represents the third contribution from the distribution along the third principal axis.

The numerical results of the three terms are shown in Fig. 8 in the Appendix. Like in Fig. 2, clear deviations of the correlation coefficients are colored based on the corresponding numerical data of the PC expansion of the inter-item-uncorrelated matrix $\mathbf{C}_{R1}$. We can see the following.

IIC.a. In $\mathbf{C}_1$ of Appendix Fig. 8(a), we find the same characteristic structure as in $\mathbf{C}$, but the deviations are larger in magnitude. This means that the characteristics [(IIB.1–3) of the last subsection] of $\mathbf{C}$ are dominated by the first term $\mathbf{C}_1$ coming from the information of the distribution of $\mathbf{s}_n (n = 1 - N)$ along the first PC axis. This is consistent with the biggest variance ($\lambda_1 = 7.282$) of the distribution along the first PC axis, showing the biggest contribution to $\mathbf{C}$ from along the first PC axis.

IIC.b1. In $\mathbf{C}_2$ of Appendix Fig. 8(b), in which the first correction term of the correlation coefficients is represented, we can find quite different features.

The correlation coefficients between the EW12 item group and EW45 group show clear negative deviation. The deviation is far greater in magnitude than the numerical fluctuation that R1 induces.

In other words, although ESJ as a whole demonstrates coherence between EW45 and EW12, there are a noteworthy minority of students contributing to the negative deviation of the correlation coefficients between the two-item groups, representing confusion or conflict between them.

IIC.b2. A similar trend is in the correlation coefficients between the EW12 group and EW3 group in $\mathbf{C}_2$, which may be related to the N2-NF and N3 confusion pointed out by Law and Wilson [45].

Statements IIC.b1 and IIC.b2 together show that the distribution is quite different in the direction of the second principal axis: some negative deviations of the correlation coefficients are observed between the item groups along the second principal axis, while positive deviations are along the first principal axis.

IIC.c1. In $\mathbf{C}_3$ of Appendix Fig. 8(c), in which the second correction term is represented, we see that, although the correlation coefficients are mostly small in magnitude, there is nevertheless a clear trend: the correlation coefficients between EW3 group and EW45 group show negative deviation. This implies that some students may experience confusion or conflict regarding the concepts of the two item groups.

IIC.c2. The correlation coefficients between EW3 group and EW12 group have a similar trend in $\mathbf{C}_3$, but some elements show weak positive deviation depending on the items of EW12 group.

Thus, we see that the distribution is quite different also in the direction of the third principal axis, with some new negative deviations and some positive deviations between the item groups appearing.

IIC.d. It is remarkable that the sum of the main three terms, $\mathbf{C}_{123} \equiv \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{C}_3$, shown in the Appendix [Fig. 8(d)], closely resembles $\mathbf{C}$. This fact suggests that the adopted three-dimensional PC space is effective enough for describing the correlation coefficients between the examinees' interitem responses.

The statement IIC.d argues that the three-dimensional PC space is effective for representing the main structure of the interitem correlation. At the same time, the statements IIC.a–c imply that the distribution of score-state vectors in the three-dimensional PC space has more internal information than that from the simple interitem correlation coefficients of the correlation matrix $\mathbf{C}$.

It is reasonable to suppose that the higher-order terms of the PC expansion mainly describe what is called "random fluctuation" of the distribution, and so we look only to the three-dimensional PC space in the next section to explore information beyond what is evident from examining the correlation matrix $\mathbf{C}$ alone.

III. NON-NORMAL DISTRIBUTION OF THE SCORE-STATE VECTORS IN PC SPACE

A. Representation of score-state vectors in PC space

From the spectral decomposition of the correlation matrix $\mathbf{C}$ in the last section, we understand that the intra- and inter-correlations in and between the three item groups have a characteristic structure depending on the direction of the distribution of score-state vectors.

However, the deviation of the interitem correlation coefficients from those of artificial inter-item-uncorrelated samples R1 reveals only partial information about the entire correlation structure of the score-state vectors of ESJ.

To observe the entire correlation structure in more detail without losing information, we have to observe the distribution in the three-dimensional PC space directly.

To represent the distribution itself, we calculate the three principal-axis components (principal components or PCs in short) of the score-state vectors. The projection of the score-state vector $\mathbf{s}_n$ to the ith principal axis (the ith PC, or i PC in short) is denoted s_n^i.

The first principal axis shows the direction along which the distribution of score-state vectors $\mathbf{s}_n (n = 1 - N)$ is the widest (the variance is λ_1) and the first eigenvector $\mathbf{v}_1$ of $\mathbf{C}$ is the only one in which all elements are of the same (positive) sign. This indicates that the first PC of the score-state vector $\mathbf{s}_n$ [denoted s_n^1 and abstractly denoted s^1 (1 PC)] is a suitably weighted sum of the item scores of the nth examinee.

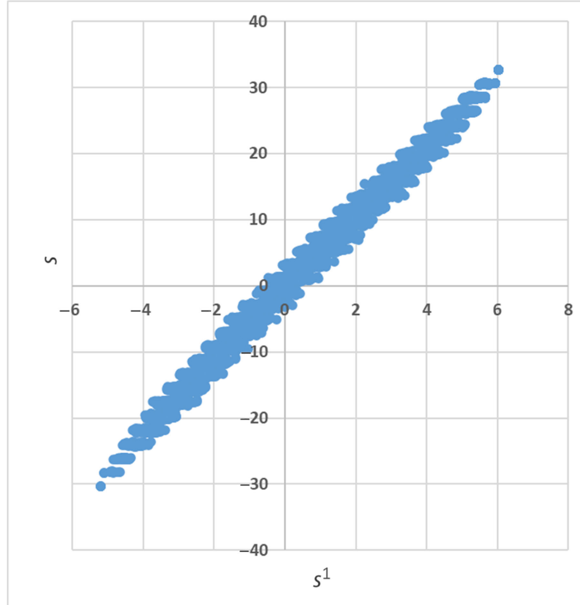


FIG. 5. Arithmetically summed total score (s) vs the first principal-axis-component (s^1) of score-state vector.

That is, s^1 suitably measures the total score and, by extension, proficiency in Newtonian mechanics.

The correlation between s^1 (which ranges from −5.199 to 6.019) and arithmetically summed total (normalized) score s (which ranges from −30.33 to 32.72) is 0.996. The correlation coefficient between s^1 and arithmetically summed total raw score s^0 also has the same value of 0.996.

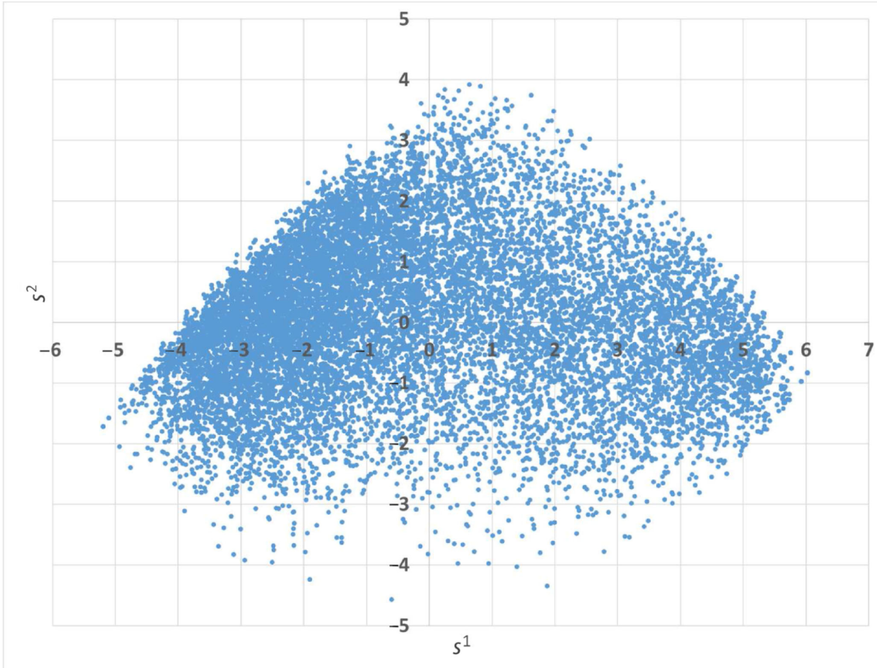
Figure 5 shows the relation between s and s^1.

A good correspondence between the arithmetically summed raw score s^0 and the proficiency in IRT has been confirmed by Wang and Bao [27]. This encourages us to use the term proficiency for s^1 in this paper.

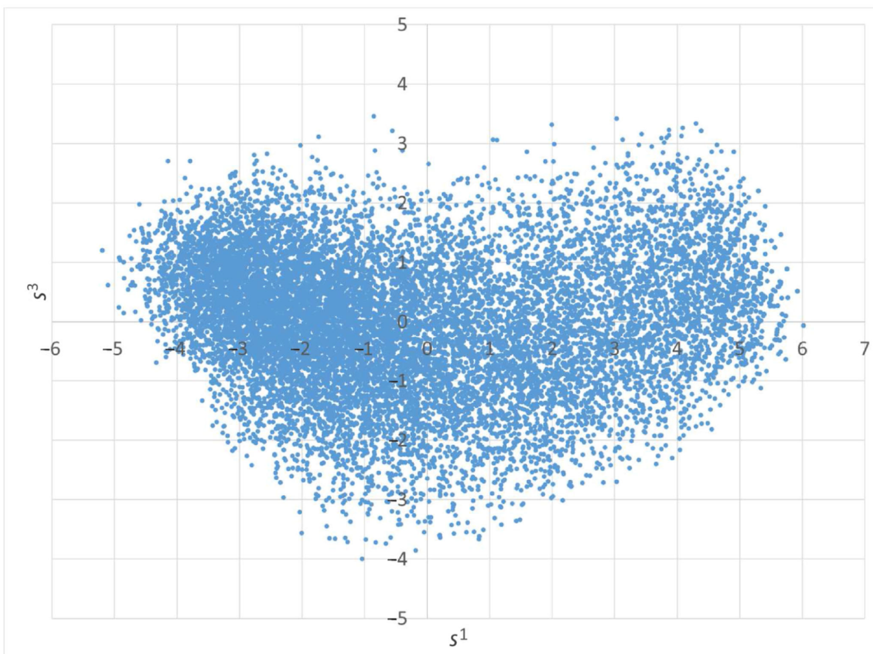
Each of the elements of the first eigenvector $\mathbf{v}_1$ represents the weight (in other words, contribution) of each item to the proficiency. From $\mathbf{v}_1$ (Eig. 1) in Table II, we can see that the $EW45$ item group on the concept of force contributes comparatively strongly to proficiency, suggesting that force is a central concept for understanding Newtonian mechanics, as pointed out by Hestenes *et al.* [1,2].

B. Non-normal distribution of the score states

In Fig. 6, the entire distribution of score-state vectors of ESJ is plotted in the space of (a) the second principal axis (vertical axis) vs the first principal axis (horizontal axis) and (b) the third principal axis vs the first principal axis. It is impressive and striking that the distribution is not symmetric in both subspaces.



(a)



(b)

FIG. 6. Distribution of the $N = 13\,768$ score-state vectors plotted (a) in the two-dimensional PC space spanned by the second PC (s^2) axis (vertical axis) vs the first PC (s^1) axis (horizontal axis), and (b) in the two-dimensional PC space spanned by the third PC (s^3) axis vs the first PC (s^1) axis. In (a), the null score-state vector is $(s^1, s^2) = (-5.119, -1.720)$, and the perfect score-state vector is $(s^1, s^2) = (6.019, -0.835)$. In (b), the null score-state vector is $(s^1, s^3) = (-5.119, 1.201)$, and the perfect score-state vector is $(s^1, s^3) = (6.019, -0.063)$.

More precisely, in Fig. 6(a), the data are densely aggregated in the left (negative) side of proficiency s^1, with a positive slope starting from the null score state $(s^1, s^2) = (-5.119, 1.720)$. The distribution then snakes downward at around $s^1 = -0.5$ and widely distributes on the positive side of proficiency s^1. The overall feature of the right side distribution gradually converges downward toward the perfect score state $(s^1, s^2) = (6.019, -0.835)$.

In Fig. 6(b), the feature of the distribution is more characteristic as if it were a snaked fat caterpillar with a fat tail, a fat body, and a small head from left to right that snakes twice toward the final goal of the right extreme of proficiency.

The dense region of data is in the negative area of proficiency, but this time with a negative slope. The fat tail part starting from the null score state $(s^1, s^3) = (-5.119, 1.201)$ directs downward to sink to the negative area of s^3 space at around the origin of s^1.

The distribution then snakes to make an upward body of the fat caterpillar and overshoots up to the positive area of s^3 space at around a value of $s^1 = 4.5$.

The small head part of the distribution then snakes again downward to correct the overshooting of the fat body toward the perfect score state $(s^1, s^3) = (6.019, -0.063)$.

As a result, in Fig. 6, both the null score state $(s^1, s^2, s^3) = (-5.199, -1.720, 1.201)$ with the lowest s^1 value of -5.119 and the perfect score state $(s^1, s^2, s^3) = (6.019, -0.835, -0.063)$ with the highest s^1 value of 6.019 lie off the first principal axis.

In contrast, the corresponding distribution of artificial score state of R1 (see Fig. 2 in Supplemental Material [44]) exhibits an ellipticlike symmetric global structure, with both the highest and lowest proficiency score state lying on the first principal axis [44].

These facts arise from the fact that the distribution of score state of ESJ is not normal, while the artificial distribution of the inter-item-uncorrelated samples of R1 is normal except along the first PC axis.

The normal (Gaussian) distribution is the one that is described by the exponential-type distribution function $\propto \exp[-f(q_1, q_2, \ldots, q_s)]$ in which the exponent $f(q_1, q_2, \ldots, q_s)$ has a quadratic form of physical quantities $q_1, q_2, \ldots,$ and q_s.

For example in the case of one dimension (with one physical quantity), the exponent is $f(q) = q^2/(2\sigma_q^2)$, and the distribution function is characterized by only one quantity σ_q; the standard deviation of the distribution that is the square root of the variance σ_q^2 (which is the second-order correlation of the distribution) As a result, all statistically averaged quantities are suitable functions of only one characteristic value of the standard deviation (and of the mean when the distribution is not centered).

In the general case of s-dimension, because of the exponent's quadratic form, the distribution's only characteristics are the standard deviations along the distribution's principal axes, which determine the covariant matrix describing the correlation of the second order. As a result, all statistically averaged quantities are suitable functions of the correlations of the second order. That is, higher-order correlations of more than the second order are described by the correlation of the second order (and of the first order, describing the mean when the distribution is not centered). This is the basis of the great advantage of the covariance analysis in the case of a normal distribution.

In cases where the distribution is non-normal (like the ESJ), however, we also need to consider correlations of orders higher than the second order because correlations of the lower orders are insufficient to describe them.

This fact gives us a warning to many previous analyses concerning correlation coefficients (covariances in general) mentioned in Sec. II A: they consider the correlation of the second order and thus only provide a perfect description of the entire correlation structure when the distribution is normal. When the distribution is not normal, mean(s) and covariance(s)—though still important—are insufficient to describe the correlation structure in the distribution.

The parallel analysis of the correlation matrices (shown in Figs. 2, 8, and 9 in this paper) compared with the reference of R1 (the inter-item-uncorrelated systems with normal distribution except along the first PC axis) considered only the correlation of the second order and so provided only partial information regarding the non-normal distribution. That is, we looked only at the correlation coefficients between two-item scores.

We need to describe higher-order correlations of more than the second order to describe the entire correlation structure.

There are many ways in statistics to describe the higher-order correlations of more than the second order.

The most natural extension from our case is the correlation of three-item scores, four-item scores, five-item scores, and so on. In that case, we need to calculate a huge number of probabilities of finding three-item clusters, four-item clusters, five-item clusters, and so on.

However, doing this step-by-step process is too lengthy and time-consuming to feasibly describe the comprehensive structure of the correlation in a transparent way.

Instead, we want to choose in this paper a primitive but strong way of using our eyesight to recognize the phase structure of highly correlated systems. Our eyes can efficiently recognize phase structures that are difficult for analytical approaches and numerical calculations to recognize.

Our aim at the present stage is to try to understand the global feature of the students' nonlinear response caused by the non-normal distribution of the score-state vectors.

Like in the study of Morris *et al.*, [34,35], we represent in this paper the entire structure of the distribution in the three-dimensional PC space without artificially simplifying or modifying the original data.

We examine the general features of the snaked non-normal distribution without predicting what function best characterizes the distribution.

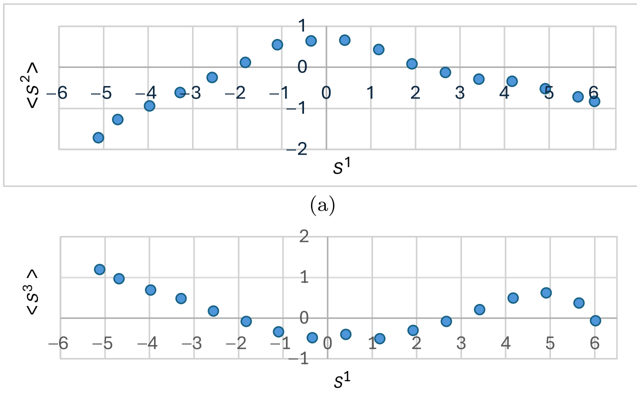


FIG. 7. (a) Local average of the second PC of score-state vector, $\langle s^2 \rangle$, as a function of s^1. (b) Local average of the third PC of score-state vector, $\langle s^3 \rangle$, as a function of s^1. In both figures, the null score-state vector and the perfect score-state vector are plotted to see the two extreme values of $\langle s^2 \rangle$ and $\langle s^3 \rangle$, respectively.

C. Novice area and transient and expert area on the first principal axis

In Figs. 6(a) and 6(b), we see a clear snaked structure in the distribution around the origin of the proficiency. And so, as a first step, we divide the distribution into two parts, a left half and a right half.

To consider where to divide the snaked structure, we calculate a local mean of the second PC of the score-state vector s^2 as a function of s^1 [see Fig. 7(a)]. The local mean of s^2, $\langle s^2(s^1) \rangle$, is calculated for the local student samples having the values of s^1 in a narrow bin suitably prepared on the s^1 axis and plotted at the local mean of the s^1 in the bin. This plot contains two extreme points: the null score-state vector and the perfect score-state vector.

The local means of s^2 calculated in all bins and at the extreme points on the s^1 axis are combined to draw the local mean $\langle s^2 \rangle$ as a function of s^1 on the first principal axis. In Fig. 7(b), the same is made for the local mean $\langle s^3 \rangle$ as a function of s^1.

We refer to these plots as the "spine" of the distribution. In the case of data that follows an inter-item-uncorrelated distribution, we would see a spine in Figs. 7(a) and 7(b) that follows the horizontal lines. With ESJ, however, we see that this is not the case.

If s^2 and s^3 do not correlate with s^1, both figures Figs. 7(a) and 7(b) should follow horizontal lines with y axis values much smaller than the standard deviations $\sigma_{2-R1} = 0.893$ and $\sigma_{3-R1} = 0.898$, respectively, of R1. We reconfirm from Figs. 7(a) and 7(b) that both the second PC of score-state vector (s^2) and the third PC of score-state vector (s^3) have snaked structures as the examinees' proficiency increases along the s^1 axis.

We can see a main snaked structure of the spine in the middle of the s^1 axis in both Figs. 7(a) and 7(b) and also see another diplike snaked structure at almost the top of the s^1 axis in Fig. 7(b).

From Fig. 7(a), the starting point of the middle snake of the spine appears to be around the value of $s^1 = -1$. From Fig. 7(b), the ending point of the middle snake of the spine appears to be around the value of $s^1 = 1$.

The narrow area between the two points in the s^1 axis suggests a transient area between the two areas on both sides that may be called the novice area and the expert area, respectively. However, it is not easy to describe numerically the narrow distribution of the transient area separately.

As a first step, we divide the distribution into two, distinguishing the novice area at the s^1 value of -0.5. The value -0.5, corresponding to the raw total-score rate between $13/30$ and $14/30$, is a qualitative threshold obtained from visual inspection and sufficient to yield this paper's qualitative results.

In the analysis that follows, we divide our data into two subensembles, $\mathrm{ESJ}_{1/2}$ (novice students with an s^1 value less than -0.5) and $\mathrm{ESJ}_{2/2}$ (transient and advanced (expert) students with an s^1 value greater than -0.5).

The division of the total distribution into two subdistributions—novice and transient and expert—presents us with a new perspective that we can always decompose the total ensemble of samples into two subensembles when the entire ensemble extends sufficiently over the entire range of the s^1 axis.

It is important to mention that novice and/or expert is not an additional independent quality given to the ensemble of samples but is rather a label distinguishing the partial areas (that is, the partial characteristic structures) of the unified global distribution in the PC space. The characteristic structure of the distribution plays an essential role in the study of this paper.

In addition to the main snake structure in both s^2 and s^3 in the distribution, we observe another snake structure in s^3, a statistical local maximum, in the upper expert area. We will discuss this again in statement IIIC.b3 of this subsection below.

For each subensemble, we calculate the eigenvalues and eigenvectors of the corresponding correlation matrices. The three main eigenvectors of the correlation matrix of $\mathrm{ESJ}_{2/2}$ (which we label $\mathbf{C}_{2/2}$) and those of $\mathrm{ESJ}_{1/2}$ ($\mathbf{C}_{1/2}$) are shown in Tables IV and V, respectively.

These two sets of the three main eigenvectors are different.

In Table VI, we compare each of the two sets of eigenvectors with those of the entire ensemble of samples ESJ by making the inner products between them.

We can check whether two eigenvectors of principal axes are close or not by checking the angle between the two eigenvectors. The angle is obtained from the inner product

TABLE IV. Three main eigenvectors (Eig.s) of the correlation matrix $\mathbf{C}_{2/2}$ describing the transient and expert part. The corresponding eigenvalues are, $\lambda_1 = 3.934$, $\lambda_2 = 1.939$, and $\lambda_3 = 1.358$.

	Eig. 1	Eig. 2	Eig. 3
1	0.1054	0.0171	−0.2856
2	0.1351	−0.0779	−0.2451
3	0.0555	0.1228	−0.0314
4	0.1343	−0.1790	−0.4420
5	0.2743	−0.2346	0.2263
6	0.1210	0.1301	−0.0768
7	0.1438	0.0297	−0.0641
8	0.1484	0.1730	0.0662
9	0.1798	0.1653	0.0458
10	0.1058	0.1025	0.0230
11	0.1827	−0.1278	0.2957
12	0.0999	0.0734	0.0198
13	0.3154	−0.1803	0.1565
14	0.1763	0.1682	0.0040
15	0.0505	−0.2131	−0.4011
16	0.1378	−0.1918	−0.0798
17	0.2355	−0.2078	−0.0608
18	0.2468	−0.2648	0.2936
19	0.1146	0.1856	−0.0392
20	0.1339	0.1705	−0.0658
21	0.1881	0.2696	0.0026
22	0.2017	0.2743	−0.0389
23	0.2027	0.3785	0.0098
24	0.1307	0.2808	−0.0548
25	0.2842	−0.1395	−0.0300
26	0.2928	−0.0451	−0.0660
27	0.1362	0.2050	−0.0414
28	0.1354	−0.0930	−0.4298
29	−0.0268	−0.1111	0.0119
30	0.3037	−0.0997	0.1485

TABLE V. Three main eigenvectors (Eig.s) of the correlation matrix $\mathbf{C}_{1/2}$ describing the novice part. The corresponding eigenvalues are, $\lambda_1 = 1.9585$, $\lambda_2 = 1.5833$, and $\lambda_3 = 1.2832$.

	Eig. 1	Eig. 2	Eig. 3
1	0.2257	−0.1758	−0.2606
2	0.0748	−0.1686	−0.3933
3	0.2114	−0.0361	0.0069
4	0.0805	−0.3936	0.0102
5	−0.2015	−0.0377	−0.0189
6	0.1438	−0.0221	−0.2797
7	0.0993	0.0282	−0.0192
8	0.1447	0.0272	0.3015
9	0.1135	0.1656	0.4400
10	0.3458	−0.1701	−0.1217
11	−0.0446	−0.0808	0.0767
12	0.1880	−0.0157	−0.0806
13	−0.1481	−0.0892	−0.1648
14	0.1450	0.0378	0.2047
15	0.0117	−0.4755	0.1499
16	0.0241	−0.4208	0.2982
17	−0.1975	−0.2418	0.1586
18	−0.1818	−0.1036	0.0342
19	0.3190	0.1194	0.1230
20	0.2762	0.0660	0.1860
21	0.1339	0.1278	0.1170
22	0.0733	0.0758	−0.0079
23	0.1233	0.1044	−0.0870
24	0.3046	−0.0254	−0.2117
25	−0.1797	−0.1675	0.1119
26	−0.0637	−0.0181	0.0521
27	0.2133	0.1598	0.0414
28	0.2463	−0.3351	−0.0066
29	0.1758	−0.1285	0.1932
30	−0.1788	−0.0332	−0.0986

of them. The smaller the angle, the closer the two eigenvectors are.

The three main eigenvectors of $\mathbf{C}_{2/2}$ of the transient and expert students are similar to those of $\mathbf{C}$ in Table II. However, the three main eigenvectors of $\mathbf{C}_{1/2}$ of novices differ significantly from those of $\mathbf{C}$.

The first eigenvector of $\mathbf{C}_{1/2}$ aligns best with the second eigenvector of $\mathbf{C}$ of the entire ensemble of samples ESJ. Similarly, the second eigenvector of $\mathbf{C}_{1/2}$ aligns best with the third eigenvector of $\mathbf{C}$ of ESJ, and the third eigenvector of the novice aligns best with the fourth eigenvector of ESJ.

While the three main principal axes of the transient and expert subensemble are similar to those of the three-dimensional ESJ ensemble, the eigenvectors of the novice subensemble are skewed like a screw.

The skewed structure of the subdistribution of novice suggests a special and delicate position of the novice states that cannot be explained by looking only at the weakness of the correlation coefficients.

The different structure of the distribution between novice area and transient and expert area has already been observed in Figs. 6(a), 6(b), 7(a), and 7(b), in the three-dimensional score-state vector space of the entire ensemble of systems.

As Hestenes *et al.* pointed out [2], the analytical result of FA is ensemble-dependent; that is, the result depends on what ensemble of samples we are dealing with.

More strictly, FA calculates the quantity, which depends on the intensity of the local distribution of the score states.

TABLE VI. Inner products between each of the main three eigenvectors (Eig.s) of novice (Nov) and transient and expert (T&E), and the four main eigenvectors (Eig.s) of the entire ensemble of samples ESJ.

	Eig. 1	Eig. 2	Eig. 3	Eig. 4
Eig.1-Nov	0.353	0.863	−0.326	0.035
Eig.2-Nov	−0.385	0.428	0.737	0.059
Eig.3-Nov	0.148	−0.047	0.132	0.715
Eig.1-T&E	0.959	−0.171	0.193	−0.068
Eig.2-T&E	0.071	0.891	0.395	−0.101
Eig.3-T&E	−0.127	−0.183	0.681	0.364

When we deal with an ensemble of samples that are primarily of novice learners, we get a different FA result from those obtained from an ensemble of samples that are primarily of expert students.

The above-mentioned characteristic structure of the non-normal distribution of students' score states of the FCI is the origin of a long-term discussion given first by Huffman and Heller [2,17,19,23]. A linear analysis of FA is a sophisticated and powerful method of statistical analysis for an ensemble of samples with normal distribution, but it is insufficient in the case of non-normal data like what is produced with the FCI.

Two correlation matrices $\mathbf{C}_{1/2}$ and $\mathbf{C}_{2/2}$ are shown in the Appendix [Figs. 9(a) and 9(b)], respectively, to describe the deviation from those of the inter-item-uncorrelated ensemble R1, though the deviation only predicts what happens within the correlation of the second order in the case of non-normal distribution.

In both figures, clear deviation of the correlation coefficients is colored based on the numerical data of the corresponding reference correlation matrices $\mathbf{C}_{1/2-R1}$ and $\mathbf{C}_{2/2-R1}$ of R1, respectively. The two reference matrices $\mathbf{C}_{1/2-R1}$ and $\mathbf{C}_{2/2-R1}$ are made from the novice part and the transient and expert part of the ensemble of the inter-item-uncorrelated random samples R1, respectively.

The averages of the uniformly distributed off-diagonal matrix elements and their standard deviations are listed in Table I in Supplemental Material [44], with those of $\mathbf{C}_{R1}$ themselves.

Appendix Fig. 9(a) shows the correlation matrix $\mathbf{C}_{1/2}$ of the novice subensemble. For novices, the correlations between items are weak, reflecting weak coherence in the students' responses to the items. However, the correlations have a clear structure, as we specify in the following notes.

IIIC.a1. The EW3 item group has a comparatively good intragroup correlation despite the weakness of the values, suggesting that novices are fairly consistent in responding to questions on the action-reaction law in a way that demonstrates naive concepts.

IIIC.a2. Correlation coefficients between the EW45 item group and EW12 are weak but show clear negative deviation from those of R1, suggesting some confusion or conflict between the individual student's understanding of the items of EW45 on the concept of force and that of the items of EW12.

A similar trend exists between the EW45 item group and EW3 item group.

Remarkably, these negative deviations are different from the correlation in Fig. 2. In Fig. 2, there are positive deviations of the correlation coefficients between EW45 and the others (EW12 and EW3). This suggests a specific situation of the minds of novice students.

IIIC.a3. The EW3 item group and EW12 item group are similarly dominated by negative deviations.

From Appendix Fig. 9(b) of $\mathbf{C}_{2/2}$, we can see the following in the transient and expert area:

IIIC.b1. Like in the entire correlation shown in Fig. 2, the EW45 item group has a good intragroup correlation, showing a large coherence in the individual student's understanding of the eight items of the EW45 item group on the concept of force.

IIIC.b2. There is a weak but positive coherence between the EW45 item group and both the EW12 and EW3 item groups, as shown in Fig. 2.

IIIC.b3. The negative deviation between the EW3 group and the EW12 group is remarkable and stronger than that in the entire correlation observed in Fig. 2. This suggests that the N2-NF and N3 confusion pointed out by Law and Wilson [45] is the most pronounced in the transient and expert area. Law and Wilson pointed out from their FCI data that as students come to understand the connection of net force to acceleration in Newton's equation of motion (N2-NF), they begin to be confused about understanding the action-reaction law (N3) in the process of comprehending the two concepts in their minds.

IV. CONCLUDING REMARKS AND DISCUSSION

This paper summarizes and extends four research reports presented in the annual and semiannual meetings of the Physical Society and Physics Education Society of Japan held from 2019 to 2021 [47–50].

In this paper, we have described the entire structure of students' states of understanding of the force concept by tracing students' correct-incorrect responses to the 30 items of the FCI without losing the structural information of the non-normal distribution of the students' score states.

We first defined a 30-dimensional score-state vector that consists of a student's correct-incorrect responses to the 30 FCI items. By projecting the score-state vectors of an assembly of $N = 13\,768$ Japanese students into a suitable low-dimensional vector space spanned by the principal axes of the correlation matrix, we found a characteristic feature of the non-normal distribution of the score-state vectors (Figs. 6 and 7).

The distribution has two snakes: a main snake and a small, diplike snake. The entire distribution is roughly divided into two subdistributions, which correspond to novice students and transient and expert students.

The first (second) eigenvector of the correlation matrix for the novices aligns best with the second (third) eigenvector of the entire ensemble of respondents, making a screwlike rotation.

Former research has shown marked differences in answering patterns on the FCI between novice and expert students. For example, factor-loading vectors of FA depend upon whether the ensemble of samples is novicelike or expertlike. At the same time, research has also shown global features of response patterns—such as IRC characteristics—that are shared between students in Japan and in the USA, as well as between high school students and university students.

We thus suspect that the "snaked" and "skewed" non-normal structure we found in Figs. 6 and 7 is a fundamental one that is independent of nationality and learning history.

When the distribution of students' score states is non-normal, we must take care in describing the entire structure of the correlation between item scores and between factors. The correlation coefficient between two physical quantities is usually useful but is not sufficient to describe the full structure of students' score states, as they form a non-normal distribution with higher-order correlations of more than the second order.

Because sophisticated analysis and modeling invariably involve losing information in extracting the essence of what is to be modeled, we must first return, as we did so in this paper, to a more primitive approach of treating the data as they are. In doing so, we observed the distribution of students' states to be snaked and skewed.

In this paper, we took two approaches to describe the higher-order correlation.

Our first approach was to use our eyes to observe the entire range of the students' score-state vectors. This allowed us to identify that the distribution is non-normal and that it is composed of two characteristic subdistributions, distinguishing novice students and expert students in the PC space. It also allowed us to find the "snaked" and "skewed" structure of the distribution.

We expect that more characteristics can be found from the direct observation of the non-normal distribution.

The principal axes were determined on the basis of the principal components analysis (PCA). We confirmed the validity of this approach by checking for a rough correspondence between PCA and CFA in the three-dimensional PC space. We also confirmed that the sum of the three main terms of the PC expansion of the correlation matrix nearly comprises the correlation matrix itself.

In addition, we observed that the PC expansion revealed the internal structure of the correlation that is not observed directly in the correlation coefficients of the entire sample systems.

We expect that these internal structures of the correlation matrix are a key to the further study of the more detailed structure of the correlation.

Our second approach of describing the higher-order correlation was to observe the deviation of the second-order correlation from that of the ideal inter-item-uncorrelated (for individuals) ensemble of samples R1.

For this purpose, we made a parallel analysis of the correlation matrices to see how our real samples with non-normal distribution compare with those of R1.

Specifically, we observed that our interitem correlation coefficients (correlation matrix) deviate from those of R1.

The main observations are in the statements IIB.1–3, IIC.a–c, and IIIC.a-b. Broadly speaking, positive deviations of the interitem correlations (indicating intra- and inter-item-group correlations) of our data from those of R1 represent coherence; negative deviations represent confusion and conflicts between ideas in the minds of individual learners.

An important position of the concept of force pointed out by Hestenes *et al.* [1,2] is observed in these statements with Table II. The EW45 item group is of central importance and is correlated with the other two-item groups, EW12 and EW3, positively and, in some cases, even negatively. Some of the negative and positive deviations of the correlations from those of R1 have been discussed extensively by Law and Wilson [45].

We anticipate that further study of the positive and negative deviations of the correlations from those of R1 will give us a deeper understanding of the students' states of understanding Newtonian mechanics.

The PCA-based representation of the score-state vector is a promising means to describe the students' states. We can extend the analysis by expanding the dimensions of the representative space to more than three. Then we can expect to realize the relation between the correlation structure of this approach and Hestenes *et al.*'s six-dimensional axes of the force concept of Newtonian mechanics, accompanied by the corresponding misconceptions [1,2], more clearly.

An advantage of the PCA-based approach is that, even when we extend the dimensions to more than three, all the results obtained in the three-dimensional PC space in this paper survive without any change as long as we use the orthonormal set of the principal axes.

In exploring the PC space, we found that the first principal axis serves as a proxy for proficiency in understanding the "force concept" of Newtonian mechanics. This suggests that we can consider the distribution to represent a statistically growing process. However, we need to clarify what the second and third principal axes describe to proceed in this direction. The factor-loading vectors obtained by Eaton and Willoughby provide us with a good suggestion [23], but we need to clarify them in the three-dimensional PC space. We will discuss this issue in a future paper.

To better understand the in-brain mechanics of student learning, it is necessary to deepen our understanding of the snaked and skewed distribution of the students' score states.

In separating the ensemble of score states into novice and intermediate and expert subpopulations, we suspect that the expert area is relatively easy to model and aligns well with previous publications in physics education research.

However, the novice area and transient area seem much more complicated and have not been studied enough [51].

In the Appendix, Fig. 9(a) shows that, in the novice area, there are negative deviations that do not appear in the transient and expert area. Furthermore, as is shown in Table VI, the score-state distribution of the novices is skewed to the distribution of the transient and expert. Furthermore, the components of the first eigenvector (the first principal axis) of the novice subdistribution do not have the same sign and seem to have some characteristic structure in them, as seen in Table V.

These suggest that the proficiency of novices, more correctly described along the first principal axis of the novice subdistribution, is not simple. The complicated structure of the distribution in the novice area cannot be understood by only the low correlation coefficients between item scores in the novice area.

ACKNOWLEDGMENTS

This study is partially supported by the Grants-in-aid for Scientific Research 26282032 of Japan Society for the Promotion of Science (JSPS). We thank Professor Shuichi Iwatsubo for his useful comments on principal component analysis and factor analysis. We also thank Dr. Keisuke Mitsumoto for introducing us to R Studio and for much help. We used R Studio to diagonalize matrices. We also thank Professor Michael M. Hull for his kind support in English and for his many useful comments on making the revised manuscript.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they are owned by a third party and the terms of use prevent public distribution. The data are available from the authors upon reasonable request.

APPENDIX: ADDITIONAL CORRELATION MATRICES

The main three terms ($\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$) of the spectral decomposition of the correlation matrix $\mathbf{C}$ of the ensemble of samples ESJ are shown in Fig. 8. Figures 8(a)–8(c) show the first, second, and third terms ($\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}_3$), respectively, of the decomposition of $\mathbf{C}$.

The correlation matrices $\mathbf{C}_{1/2}$ of novices and $\mathbf{C}_{2/2}$ of transient and expert of ESJ are shown in Fig. 9. They are made by dividing the entire ensemble of score-state vectors of ESJ into two, depending on whether the first principal component of the score-state vector is less than -0.5 or more than -0.5, respectively. The value of -0.5 corresponds to the raw total-score rate of about $14/30$.

The corresponding two inter-item-uncorrelated samples are made by dividing R1 into two, keeping the total-score distributions of novice and transient and expert, respectively, the same as those of $\mathbf{C}_{1/2}$ and $\mathbf{C}_{2/2}$.

1st term		EW12 Item Group														EW3 Item Group				EW45 Item Group							
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26
	Q6	0.12	0.15	0.18	0.18	0.18	0.18	0.17	0.18	0.14	0.19	0.16	0.13	0.17	0.14	0.16	0.11	0.19	0.18	0.19	0.19	0.24	0.19	0.22	0.18	0.2	0.21
	Q7	0.15	0.18	0.22	0.22	0.22	0.22	0.21	0.22	0.18	0.23	0.2	0.16	0.21	0.18	0.19	0.14	0.24	0.22	0.23	0.23	0.3	0.24	0.27	0.22	0.25	0.26
	Q8	0.18	0.22	0.26	0.27	0.27	0.27	0.25	0.26	0.21	0.28	0.24	0.19	0.26	0.21	0.23	0.17	0.29	0.27	0.28	0.28	0.36	0.29	0.32	0.26	0.31	0.31
	Q10	0.18	0.22	0.27	0.28	0.28	0.28	0.26	0.27	0.22	0.29	0.25	0.2	0.26	0.22	0.24	0.17	0.29	0.28	0.29	0.29	0.37	0.3	0.33	0.27	0.31	0.32
	Q20	0.18	0.22	0.27	0.28	0.27	0.27	0.26	0.27	0.22	0.29	0.25	0.19	0.26	0.22	0.24	0.17	0.29	0.27	0.28	0.29	0.36	0.3	0.33	0.27	0.31	0.32
	Q23	0.18	0.22	0.27	0.28	0.27	0.27	0.26	0.27	0.22	0.29	0.25	0.19	0.26	0.22	0.24	0.17	0.29	0.27	0.29	0.29	0.36	0.3	0.33	0.27	0.31	0.32
EW12	Q24	0.17	0.21	0.25	0.26	0.26	0.26	0.24	0.25	0.2	0.27	0.23	0.18	0.24	0.2	0.22	0.16	0.27	0.26	0.27	0.27	0.34	0.28	0.31	0.25	0.29	0.3
	Q9	0.18	0.22	0.26	0.27	0.27	0.27	0.25	0.26	0.21	0.28	0.24	0.19	0.25	0.21	0.23	0.17	0.29	0.27	0.28	0.28	0.35	0.29	0.32	0.26	0.3	0.31
	Q12	0.14	0.18	0.21	0.22	0.22	0.22	0.2	0.21	0.17	0.23	0.2	0.15	0.21	0.17	0.19	0.14	0.23	0.22	0.23	0.23	0.29	0.23	0.26	0.21	0.25	0.25
	Q14	0.19	0.23	0.28	0.29	0.29	0.29	0.27	0.28	0.23	0.3	0.26	0.2	0.27	0.23	0.25	0.18	0.31	0.29	0.3	0.3	0.38	0.31	0.34	0.28	0.33	0.34
	Q19	0.16	0.2	0.24	0.25	0.25	0.25	0.23	0.24	0.2	0.26	0.23	0.18	0.24	0.2	0.22	0.16	0.27	0.25	0.26	0.26	0.33	0.27	0.3	0.24	0.28	0.29
	Q21	0.13	0.16	0.19	0.2	0.19	0.19	0.18	0.19	0.15	0.2	0.18	0.14	0.18	0.15	0.17	0.12	0.21	0.19	0.2	0.21	0.26	0.21	0.23	0.19	0.22	0.23
	Q22	0.17	0.21	0.26	0.26	0.26	0.26	0.24	0.25	0.21	0.27	0.24	0.18	0.25	0.21	0.22	0.16	0.28	0.26	0.27	0.27	0.35	0.28	0.31	0.25	0.3	0.3
	Q27	0.14	0.18	0.21	0.22	0.22	0.22	0.2	0.21	0.17	0.23	0.2	0.15	0.21	0.17	0.19	0.14	0.23	0.22	0.23	0.23	0.29	0.24	0.26	0.21	0.25	0.25
	Q4	0.16	0.19	0.23	0.24	0.24	0.24	0.22	0.23	0.19	0.25	0.22	0.17	0.22	0.19	0.2	0.15	0.25	0.24	0.25	0.25	0.31	0.25	0.28	0.23	0.27	0.28
EW3	Q15	0.11	0.14	0.17	0.17	0.17	0.17	0.16	0.17	0.14	0.18	0.16	0.12	0.16	0.14	0.15	0.11	0.19	0.17	0.18	0.18	0.23	0.19	0.21	0.17	0.2	0.2
	Q16	0.19	0.24	0.29	0.29	0.29	0.29	0.27	0.29	0.23	0.31	0.27	0.21	0.28	0.23	0.25	0.19	0.31	0.29	0.3	0.31	0.39	0.32	0.35	0.29	0.33	0.34
	Q28	0.18	0.22	0.27	0.28	0.27	0.27	0.26	0.27	0.22	0.29	0.25	0.19	0.26	0.22	0.24	0.17	0.29	0.27	0.29	0.29	0.36	0.3	0.33	0.27	0.31	0.32
	Q5	0.19	0.23	0.28	0.29	0.28	0.29	0.27	0.28	0.23	0.3	0.26	0.2	0.27	0.23	0.25	0.18	0.3	0.29	0.3	0.3	0.38	0.31	0.34	0.28	0.32	0.33
	Q11	0.19	0.23	0.28	0.29	0.29	0.29	0.27	0.28	0.23	0.3	0.26	0.21	0.27	0.23	0.25	0.18	0.31	0.29	0.3	0.31	0.38	0.31	0.35	0.28	0.33	0.34
	Q13	0.24	0.3	0.36	0.37	0.36	0.36	0.34	0.35	0.29	0.38	0.33	0.26	0.35	0.29	0.31	0.23	0.39	0.36	0.38	0.38	0.48	0.39	0.44	0.36	0.41	0.42
EW45	Q18	0.19	0.24	0.29	0.3	0.3	0.3	0.28	0.29	0.23	0.31	0.27	0.21	0.28	0.24	0.25	0.19	0.32	0.3	0.31	0.31	0.39	0.32	0.35	0.29	0.34	0.34
	Q30	0.22	0.27	0.32	0.33	0.33	0.33	0.31	0.32	0.26	0.34	0.3	0.23	0.31	0.26	0.28	0.21	0.35	0.33	0.34	0.35	0.44	0.35	0.39	0.32	0.37	0.38
	Q17	0.18	0.22	0.26	0.27	0.27	0.27	0.25	0.26	0.21	0.28	0.24	0.19	0.25	0.21	0.23	0.17	0.29	0.27	0.28	0.28	0.36	0.29	0.32	0.26	0.3	0.31
	Q25	0.2	0.25	0.31	0.31	0.31	0.31	0.29	0.3	0.25	0.33	0.28	0.22	0.3	0.25	0.27	0.2	0.33	0.31	0.32	0.33	0.41	0.34	0.37	0.3	0.35	0.36
	Q26	0.21	0.26	0.31	0.32	0.32	0.32	0.3	0.31	0.25	0.34	0.29	0.23	0.3	0.25	0.28	0.2	0.34	0.32	0.33	0.34	0.42	0.34	0.38	0.31	0.36	0.37

(a)

2nd term		EW12 Item Group														EW3 Item Group				EW45 Item Group							
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26
	Q6	0.04	0.01	0.03	0.05	0.04	0.05	0.06	0.03	0.04	0.02	0.06	0.04	0.03	0.05	-0	-0	-0	0.01	-0.1	-0	-0	-0.1	-0	-0.1	-0.1	-0
	Q7	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.01	0.02	0.01	0.02	0.02	0.01	0.02	-0	-0	-0	0	-0	-0	-0	-0	-0	-0	-0	-0
	Q8	0.03	0.01	0.03	0.05	0.04	0.05	0.06	0.03	0.03	0.02	0.05	0.04	0.03	0.05	-0	-0	-0	0.01	-0.1	-0	-0	-0.1	-0	-0.1	-0.1	-0
	Q10	0.05	0.02	0.05	0.07	0.06	0.07	0.09	0.04	0.05	0.04	0.08	0.06	0.04	0.07	-0	-0.1	-0	0.01	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
	Q20	0.04	0.02	0.04	0.06	0.06	0.06	0.08	0.03	0.05	0.03	0.07	0.05	0.04	0.07	-0	-0	-0	0.01	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0
	Q23	0.05	0.02	0.05	0.07	0.06	0.07	0.09	0.04	0.05	0.03	0.08	0.06	0.04	0.07	-0	-0.1	-0	0.01	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
EW12	Q24	0.06	0.03	0.06	0.09	0.08	0.09	0.1	0.05	0.06	0.04	0.1	0.07	0.05	0.09	-0.1	-0.1	-0	0.02	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
	Q9	0.03	0.01	0.03	0.04	0.03	0.04	0.05	0.02	0.03	0.02	0.04	0.03	0.02	0.04	-0	-0	-0	0.01	-0.1	-0	-0	-0.1	-0	-0.1	-0	-0
	Q12	0.04	0.02	0.03	0.05	0.05	0.05	0.06	0.03	0.04	0.03	0.06	0.04	0.03	0.05	-0	-0	-0	0.01	-0.1	-0	-0.1	-0.1	-0	-0.1	-0.1	-0
	Q14	0.02	0.01	0.02	0.04	0.03	0.03	0.04	0.02	0.03	0.02	0.04	0.03	0.02	0.04	-0	-0	-0	0.01	-0.1	-0	-0	-0	-0	-0	-0	-0
	Q19	0.06	0.02	0.05	0.08	0.07	0.08	0.1	0.04	0.06	0.04	0.09	0.06	0.05	0.08	-0	-0.1	-0	0.02	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
	Q21	0.04	0.02	0.04	0.06	0.05	0.06	0.07	0.03	0.04	0.03	0.06	0.05	0.03	0.06	-0	-0	-0	0.01	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0
	Q22	0.03	0.01	0.03	0.04	0.04	0.04	0.05	0.02	0.03	0.02	0.05	0.03	0.03	0.04	-0	-0	-0	0.01	-0.1	-0	-0	-0.1	-0	-0.1	-0	-0
	Q27	0.05	0.02	0.05	0.07	0.07	0.07	0.09	0.04	0.05	0.04	0.08	0.06	0.04	0.07	-0	-0.1	-0	0.01	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
	Q4	-0	-0	-0	-0	-0	-0	-0.1	-0	-0	-0	-0	-0	-0	-0	0.02	0.03	0.02	-0	0.06	0.03	0.04	0.06	0.04	0.06	0.05	0.03
EW3	Q15	-0	-0	-0	-0.1	-0	-0.1	-0.1	-0	-0	-0	-0.1	-0	-0	-0.1	0.03	0.04	0.03	-0	0.07	0.03	0.05	0.07	0.05	0.07	0.06	0.04
	Q16	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	0.02	0.03	0.02	-0	0.06	0.02	0.04	0.05	0.04	0.05	0.04	0.03
	Q28	0.01	0	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	-0	-0	-0	0	-0	-0	-0	-0	-0	-0	-0	-0
	Q5	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.06	0.07	0.06	-0	0.15	0.06	0.1	0.14	0.1	0.14	0.12	0.08
	Q11	-0	-0	-0	-0	-0	-0	-0.1	-0	-0	-0	-0.1	-0	-0	-0	0.03	0.03	0.02	-0	0.06	0.03	0.04	0.06	0.04	0.06	0.05	0.03
	Q13	-0	-0	-0	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0	-0.1	-0.1	-0	-0.1	0.04	0.05	0.04	-0	0.1	0.04	0.07	0.1	0.07	0.1	0.08	0.05
EW45	Q18	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.1	0.06	0.07	0.05	-0	0.14	0.06	0.1	0.13	0.09	0.13	0.11	0.07
	Q30	-0	-0	-0	-0.1	-0.1	-0.1	-0.1	-0	-0	-0	-0.1	-0.1	-0	-0.1	0.04	0.05	0.04	-0	0.1	0.04	0.07	0.09	0.07	0.1	0.08	0.05
	Q17	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.1	0.06	0.07	0.05	-0	0.14	0.06	0.1	0.13	0.1	0.14	0.11	0.07
	Q25	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0	-0.1	-0.1	-0	-0.1	0.05	0.06	0.04	-0	0.12	0.05	0.08	0.11	0.08	0.11	0.09	0.06
	Q26	-0	-0	-0	-0.1	-0	-0.1	-0.1	-0	-0	-0	-0.1	-0	-0	-0.1	0.03	0.04	0.03	-0	0.08	0.03	0.05	0.07	0.05	0.07	0.06	0.04

(b)

FIG. 8. (Continued)

3rd term		EW12 Item Group														EW3 Item Group				EW45 Item Group							
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26
EW12	Q6	0	0	-0	0.01	0	-0	0	-0	0.01	-0	0	-0	-0	-0	0.02	0.03	0.01	0.02	-0	-0	-0	-0	-0	-0	-0	-0
	Q7	0	0	-0	0.01	0	-0	0	-0	0	-0	0	-0	-0	-0	0.02	0.02	0.01	0.02	-0	-0	-0	-0	-0	-0	-0	-0
	Q8	-0	-0	0	-0	-0	0.01	-0	0	-0	0	-0	0.01	0.01	0	-0	-0	-0	-0	0	0	0	0	0	0	0	0
	Q10	0.01	0.01	-0	0.05	0	-0.1	0	-0	0.02	-0	0	-0.1	-0.1	-0	0.09	0.12	0.04	0.09	-0	-0	-0	-0	-0	-0	-0	-0
	Q20	0	0	-0	0	0	-0	0	-0	0	-0	0	-0	-0	-0	0	0.01	0	0	-0	-0	-0	-0	-0	-0	-0	-0
	Q23	-0	-0	0.01	-0.1	-0	0.07	-0	0.04	-0	0.04	-0	0.09	0.07	0.03	-0.1	-0.1	-0.1	-0.1	0.04	0.02	0.02	0.02	0.05	0.01	0.04	0.05
	Q24	0	0	-0	0	0	-0	0	-0	0	-0	0	-0	-0	-0	0	0	0	0	-0	-0	-0	-0	-0	-0	-0	-0
	Q9	-0	-0	0	-0	-0	0.04	-0	0.02	-0	0.02	-0	0.05	0.04	0.02	-0.1	-0.1	-0	-0.1	0.02	0.01	0.01	0.01	0.03	0	0.02	0.03
	Q12	0.01	0	-0	0.02	0	-0	0	-0	0.01	-0	0	-0	-0	-0	0.04	0.05	0.02	0.04	-0	-0	-0	-0	-0	-0	-0	-0
	Q14	-0	-0	0	-0	-0	0.04	-0	0.02	-0	0.02	-0	0.04	0.03	0.02	-0.1	-0.1	-0	-0.1	0.02	0.01	0.01	0.01	0.02	0	0.02	0.03
	Q19	0	0	-0	0	0	-0	0	-0	0	-0	0	-0	-0	-0	0	0	0	0	-0	-0	-0	-0	-0	-0	-0	-0
	Q21	-0	-0	0.01	-0.1	-0	0.09	-0	0.05	-0	0.04	-0	0.1	0.08	0.04	-0.1	-0.2	-0.1	-0.1	0.04	0.02	0.03	0.03	0.06	0.01	0.04	0.06
	Q22	-0	-0	0.01	-0.1	-0	0.07	-0	0.04	-0	0.03	-0	0.08	0.06	0.03	-0.1	-0.1	-0	-0.1	0.03	0.01	0.02	0.02	0.05	0.01	0.03	0.05
	Q27	-0	-0	0	-0	-0	0.03	-0	0.02	-0	0.02	-0	0.04	0.03	0.01	-0	-0.1	-0	-0	0.01	0.01	0.01	0.01	0.02	0	0.02	0.02
EW3	Q4	0.02	0.02	-0	0.09	0	-0.1	0	-0.1	0.04	-0.1	0	-0.1	-0.1	-0	0.15	0.19	0.07	0.14	-0.1	-0	-0	-0	-0.1	-0	-0.1	-0.1
	Q15	0.03	0.02	-0	0.12	0.01	-0.1	0	-0.1	0.05	-0.1	0	-0.2	-0.1	-0.1	0.19	0.25	0.1	0.19	-0.1	-0	-0	-0	-0.1	-0	-0.1	-0.1
	Q16	0.01	0.01	-0	0.04	0	-0.1	0	-0	0.02	-0	0	-0.1	-0	-0	0.07	0.1	0.04	0.07	-0	-0	-0	-0	-0	-0	-0	-0
	Q28	0.02	0.02	-0	0.09	0	-0.1	0	-0.1	0.04	-0.1	0	-0.1	-0.1	-0	0.14	0.19	0.07	0.14	-0	-0	-0	-0	-0.1	-0	-0	-0.1
EW45	Q5	-0	-0	0	-0	-0	0.04	-0	0.02	-0	0.02	-0	0.04	0.03	0.01	-0.1	-0.1	-0	-0	0.02	0.01	0.01	0.01	0.02	0	0.02	0.02
	Q11	-0	-0	0	-0	-0	0.02	-0	0.01	-0	0.01	-0	0.02	0.01	0.01	-0	-0	-0	-0	0.01	0	0	0	0.01	0	0.01	0.01
	Q13	-0	-0	0	-0	-0	0.02	-0	0.01	-0	0.01	-0	0.03	0.02	0.01	-0	-0	-0	-0	0.01	0	0.01	0.01	0.01	0	0.01	0.01
	Q18	-0	-0	0	-0	-0	0.02	-0	0.01	-0	0.01	-0	0.03	0.02	0.01	-0	-0	-0	-0	0.01	0	0.01	0.01	0.02	0	0.01	0.02
	Q30	-0	-0	0	-0	-0	0.05	-0	0.03	-0	0.02	-0	0.06	0.05	0.02	-0.1	-0.1	-0	-0.1	0.02	0.01	0.01	0.02	0.03	0.01	0.02	0.03
	Q17	-0	-0	0	-0	-0	0.01	-0	0	-0	0	-0	0.01	0.01	0	-0	-0	-0	-0	0	0	0	0	0.01	0	0	0.01
	Q25	-0	-0	0	-0	-0	0.04	-0	0.02	-0	0.02	-0	0.04	0.03	0.02	-0.1	-0.1	-0	-0	0.02	0.01	0.01	0.01	0.02	0	0.02	0.02
	Q26	-0	-0	0	-0	-0	0.05	-0	0.03	-0	0.03	-0	0.06	0.05	0.02	-0.1	-0.1	-0	-0.1	0.02	0.01	0.01	0.02	0.03	0.01	0.02	0.03

(c)

Sum of the three terms		EW12 Item Group														EW3 Item Group				EW45 Item Group								EWe			
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26	Q1	Q2	Q3	Q29
EW12	Q6	0.16	0.16	0.21	0.25	0.23	0.21	0.23	0.19	0.19	0.21	0.22	0.15	0.19	0.19	0.15	0.11	0.18	0.21	0.11	0.16	0.18	0.12	0.16	0.1	0.14	0.16	0.2	0.12	0.17	0.08
	Q7	0.16	0.19	0.23	0.25	0.24	0.23	0.23	0.22	0.2	0.24	0.23	0.16	0.21	0.19	0.2	0.15	0.23	0.24	0.2	0.22	0.27	0.21	0.24	0.19	0.22	0.24	0.21	0.15	0.16	0.07
	Q8	0.21	0.23	0.3	0.31	0.31	0.32	0.31	0.29	0.24	0.31	0.3	0.24	0.29	0.27	0.19	0.12	0.26	0.27	0.21	0.26	0.31	0.23	0.28	0.2	0.25	0.28	0.25	0.16	0.21	0.07
	Q10	0.25	0.25	0.31	0.4	0.34	0.28	0.35	0.27	0.29	0.29	0.33	0.18	0.25	0.27	0.29	0.24	0.3	0.38	0.15	0.23	0.28	0.18	0.22	0.16	0.2	0.23	0.33	0.23	0.26	0.16
	Q20	0.23	0.24	0.31	0.34	0.33	0.33	0.33	0.3	0.26	0.32	0.32	0.24	0.3	0.28	0.2	0.13	0.26	0.29	0.19	0.25	0.3	0.21	0.26	0.18	0.24	0.27	0.27	0.17	0.23	0.09
	Q23	0.21	0.23	0.32	0.28	0.33	0.42	0.34	0.34	0.24	0.36	0.33	0.34	0.37	0.32	0.09	-0	0.2	0.19	0.22	0.26	0.32	0.22	0.31	0.18	0.27	0.32	0.2	0.09	0.22	0
	Q24	0.23	0.23	0.31	0.35	0.33	0.34	0.34	0.29	0.27	0.31	0.33	0.25	0.29	0.29	0.17	0.1	0.23	0.27	0.14	0.22	0.25	0.16	0.22	0.13	0.19	0.23	0.27	0.15	0.24	0.09
	Q9	0.19	0.22	0.29	0.27	0.3	0.34	0.29	0.3	0.22	0.32	0.29	0.27	0.31	0.27	0.15	0.07	0.24	0.22	0.24	0.27	0.33	0.25	0.31	0.21	0.28	0.31	0.21	0.13	0.19	0.03
	Q12	0.19	0.2	0.24	0.29	0.26	0.24	0.27	0.22	0.22	0.24	0.26	0.16	0.21	0.21	0.2	0.15	0.22	0.26	0.14	0.19	0.23	0.16	0.19	0.14	0.17	0.2	0.24	0.16	0.19	0.1
	Q14	0.21	0.24	0.31	0.29	0.32	0.36	0.31	0.32	0.24	0.34	0.3	0.28	0.33	0.28	0.17	0.09	0.26	0.24	0.27	0.29	0.36	0.27	0.33	0.24	0.3	0.33	0.23	0.14	0.21	0.04
	Q19	0.22	0.23	0.3	0.33	0.32	0.33	0.33	0.29	0.26	0.3	0.32	0.24	0.29	0.28	0.17	0.1	0.22	0.27	0.14	0.21	0.25	0.16	0.22	0.13	0.19	0.23	0.26	0.15	0.23	0.08
	Q21	0.15	0.16	0.24	0.18	0.24	0.34	0.25	0.27	0.16	0.28	0.24	0.28	0.3	0.25	0.01	-0.1	0.12	0.09	0.16	0.19	0.23	0.16	0.23	0.12	0.2	0.24	0.12	0.03	0.15	-0
	Q22	0.19	0.21	0.29	0.25	0.3	0.37	0.29	0.31	0.21	0.33	0.29	0.3	0.34	0.28	0.1	0.01	0.21	0.17	0.24	0.26	0.32	0.24	0.32	0.2	0.28	0.32	0.18	0.09	0.19	-0
	Q27	0.19	0.19	0.27	0.27	0.28	0.32	0.29	0.27	0.21	0.28	0.28	0.25	0.28	0.26	0.1	0.03	0.17	0.19	0.14	0.19	0.23	0.15	0.21	0.12	0.18	0.22	0.2	0.1	0.2	0.04
EW3	Q4	0.15	0.2	0.19	0.29	0.2	0.09	0.17	0.15	0.2	0.17	0.17	0.01	0.1	0.1	0.38	0.37	0.35	0.37	0.26	0.25	0.32	0.28	0.25	0.27	0.26	0.24	0.28	0.26	0.14	0.16
	Q15	0.11	0.15	0.12	0.24	0.13	-0	0.1	0.07	0.15	0.09	0.1	-0.1	0.01	0.03	0.37	0.4	0.31	0.35	0.19	0.19	0.24	0.21	0.17	0.22	0.19	0.15	0.25	0.26	0.1	0.18
	Q16	0.18	0.23	0.26	0.3	0.26	0.2	0.23	0.24	0.22	0.26	0.22	0.12	0.21	0.17	0.35	0.31	0.37	0.36	0.34	0.32	0.41	0.35	0.35	0.33	0.35	0.33	0.28	0.25	0.17	0.12
	Q28	0.21	0.24	0.27	0.38	0.29	0.19	0.27	0.22	0.26	0.24	0.27	0.09	0.17	0.19	0.37	0.35	0.36	0.41	0.22	0.26	0.32	0.25	0.25	0.24	0.25	0.24	0.34	0.27	0.22	0.19
EW45	Q5	0.11	0.2	0.21	0.15	0.19	0.22	0.14	0.24	0.14	0.27	0.14	0.16	0.24	0.14	0.26	0.19	0.34	0.22	0.46	0.37	0.49	0.46	0.47	0.43	0.46	0.43	0.16	0.17	0.08	-0
	Q11	0.16	0.22	0.26	0.23	0.25	0.26	0.22	0.27	0.19	0.29	0.21	0.19	0.26	0.19	0.25	0.19	0.32	0.26	0.37	0.34	0.43	0.38	0.4	0.35	0.39	0.38	0.21	0.18	0.14	0.04
	Q13	0.18	0.27	0.31	0.28	0.3	0.32	0.25	0.33	0.23	0.36	0.25	0.23	0.32	0.23	0.32	0.24	0.41	0.32	0.49	0.43	0.56	0.5	0.52	0.46	0.51	0.49	0.26	0.23	0.16	0.04
	Q18	0.12	0.21	0.23	0.18	0.21	0.22	0.16	0.25	0.16	0.27	0.16	0.16	0.24	0.15	0.28	0.21	0.35	0.25	0.46	0.38	0.5	0.46	0.46	0.43	0.46	0.43	0.18	0.18	0.09	0.01
	Q30	0.16	0.24	0.28	0.22	0.26	0.31	0.22	0.31	0.19	0.33	0.22	0.23	0.32	0.21	0.25	0.17	0.35	0.25	0.47	0.4	0.52	0.46	0.49	0.42	0.48	0.47	0.2	0.18	0.13	-0
	Q17	0.1	0.19	0.2	0.16	0.18	0.18	0.13	0.21	0.14	0.24	0.13	0.12	0.2	0.12	0.27	0.22	0.33	0.24	0.43	0.35	0.46	0.43	0.42	0.4	0.42	0.39	0.17	0.18	0.08	0.02
	Q25	0.14	0.22	0.25	0.2	0.24	0.27	0.19	0.28	0.17	0.3	0.19	0.2	0.28	0.18	0.26	0.19	0.35	0.25	0.46	0.39	0.51	0.46	0.48	0.42	0.46	0.45	0.19	0.18	0.12	0.01
	Q26	0.16	0.24	0.28	0.23	0.27	0.32	0.23	0.31	0.2	0.33	0.23	0.24	0.32	0.22	0.24	0.15	0.33	0.24	0.43	0.38	0.49	0.43	0.47	0.39	0.45	0.45	0.2	0.17	0.14	-0
EWe	Q1	0.2	0.21	0.25	0.33	0.27	0.2	0.27	0.21	0.24	0.23	0.26	0.12	0.18	0.2	0.28	0.25	0.28	0.34	0.16	0.21	0.26	0.18	0.2	0.17	0.19	0.2	0.29	0.22	0.21	0.15
	Q2	0.12	0.15	0.16	0.23	0.17	0.09	0.15	0.13	0.16	0.14	0.15	0.03	0.09	0.1	0.26	0.26	0.25	0.27	0.17	0.18	0.23	0.18	0.18	0.18	0.18	0.17	0.22	0.19	0.12	0.12
	Q3	0.17	0.16	0.21	0.26	0.23	0.22	0.24	0.19	0.19	0.21	0.23	0.15	0.19	0.2	0.14	0.1	0.17	0.22	0.08	0.14	0.16	0.09	0.13	0.08	0.12	0.14	0.21	0.12	0.18	0.09
	Q29	0.08	0.07	0.07	0.16	0.09	0	0.09	0.03	0.1	0.04	0.08	-0	-0	0.04	0.16	0.18	0.12	0.19	-0	0.04	0.04	0.01	-0	0.02	0.01	-0	0.15	0.12	0.09	0.12

(d)

FIG. 8. (a) First term, (b) second term, (c) third term, and (d) sum of the three terms, of the PC expansion of the correlation matrix C. The order of item (Q) number has been arranged to classify the three item groups EW12, EW3, and EW45, and the other four items of EW*e* are omitted except for (d). The warm colors are put on the correlation coefficients that deviate from the mean value of the corresponding PC expansion of $\mathbf{C}_{R1}$ ($\mathbf{C}_{1-R1}$, $\mathbf{C}_{2-R1}$, or $\mathbf{C}_{3-R1}$) or $\mathbf{C}_{R1}$ itself positively more than [in (a), (b), and (d), more than double of] the standard deviation of the corresponding PC expansion of $\mathbf{C}_{R1}$ or $\mathbf{C}_{R1}$ itself. The cool colors with lateral stripes are put on them that deviate from the corresponding mean value mentioned above negatively more than [in (a), (b), and (d), more than double of] the corresponding standard deviation, respectively.

Novice		EW12 Item Group														EW3 Item Group				EW45 Item Group							
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26
	Q6	1	0.07	-0	0.09	-0	0.04	0.06	-0	0.07	-0	-0	-0	0.01	0.03	0	0	-0	0.04	-0.1	-0	-0	-0	-0	-0	-0	-0
	Q7	0.07	1	0.02	0.03	0.02	0.02	-0	0.02	0.04	0.02	0.01	0.02	0.02	-0	-0	0	0	-0	-0	-0	-0	-0	-0	-0	-0	-0
	Q8	-0	0.02	1	0.07	0.05	0.07	0.02	0.15	0.06	0.04	0.01	-0	-0	0.06	0.01	0.01	0.01	0.01	-0	0.02	-0	0	-0	-0	-0	-0
	Q10	0.09	0.03	0.07	1	0.09	0.03	0.18	-0	0.11	0.05	0.08	0	0.02	0.07	0.04	0.04	0.07	0.13	-0.1	0.04	-0	-0.1	-0	-0.1	-0.1	-0
	Q20	-0	0.02	0.05	0.09	1	-0	0.05	0.06	0.02	0.03	0.25	0.04	0.01	0.05	0.02	-0	-0	0.06	-0.1	0	-0	-0	-0	-0.1	-0.1	-0
	Q23	0.04	0.02	0.07	0.03	-0	1	0.17	0	0.04	0.06	-0	0.04	0.03	0.03	-0	-0	-0	-0	-0	-0	-0	-0	0.01	-0	-0	0
EW12	Q24	0.06	-0	0.02	0.18	0.05	0.17	1	-0	0.06	0.02	0.08	0.02	0.03	0.07	-0	-0	-0	0.11	-0.1	-0	-0	-0.1	-0.1	-0	-0	-0
	Q9	-0	0.02	0.15	-0	0.06	0	-0	1	0.01	0.07	0.07	0.04	-0	0.03	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	-0	0.01
	Q12	0.07	0.04	0.06	0.11	0.02	0.04	0.06	0.01	1	0.03	0.03	0.02	0.02	0.05	0.01	0.01	0	0.04	-0	-0	0	-0	-0	-0	-0	-0
	Q14	-0	0.02	0.04	0.05	0.03	0.06	0.02	0.07	0.03	1	0.07	0.04	0.04	0.05	0.02	0.01	0.02	0.02	-0	-0	-0	0	-0	-0	-0	0
	Q19	-0	0.01	0.01	0.08	0.25	-0	0.08	0.07	0.03	0.07	1	0.06	0.02	0.08	-0	-0	-0	0.06	-0.1	-0	-0	-0.1	-0.1	-0.1	-0.1	-0
	Q21	-0	0.02	-0	0	0.04	0.04	0.02	0.04	0.02	0.04	0.06	1	0.03	0.04	-0	-0	0.02	0.02	-0	-0	-0	-0	-0	-0.1	-0	-0
	Q22	0.01	0.02	-0	0.02	0.01	0.03	0.03	-0	0.02	0.04	0.02	0.03	1	0.01	-0	-0	0	-0	-0	-0	-0	-0	-0	-0	-0	0.05
	Q27	0.03	-0	0.06	0.07	0.05	0.03	0.07	0.03	0.05	0.05	0.08	0.04	0.01	1	-0	-0.1	-0	0.03	-0	-0	-0	-0	-0	-0.1	-0.1	-0
	Q4	0	-0	0.01	0.04	0.02	-0	-0	-0	0.01	0.02	-0	-0	-0	-0	1	0.16	0.08	0.19	-0	0.02	0.01	-0	0	0.02	0.01	0.02
EW3	Q15	0	0	0.01	0.04	-0	-0	-0	-0	0.01	0.01	-0	-0	-0	-0.1	0.16	1	0.23	0.12	-0	0	0.03	0.03	0	0.09	0.03	-0
	Q16	-0	0	0.01	0.07	-0	-0	-0	-0	0	0.02	-0	0.02	0	-0	0.08	0.23	1	0.08	0	0.02	-0	0.03	-0	0.12	0.08	-0
	Q28	0.04	-0	0.01	0.13	0.06	-0	0.11	-0	0.04	0.02	0.06	0.02	-0	0.03	0.19	0.12	0.08	1	-0	-0	-0	-0	-0	-0	-0	-0
	Q5	-0.1	-0	-0	-0.1	-0.1	-0	-0.1	-0	-0	-0	-0.1	-0	-0	-0	-0	-0	0	-0	1	0.05	0.06	0.12	0.02	0.03	0.01	0.01
	Q11	-0	-0	0.02	0.04	0	-0	-0	-0	-0	-0	-0	-0	-0	-0	0.02	0	0.02	-0	0.05	1	-0	0.07	0.04	-0	0.01	0.01
	Q13	-0	-0	-0	-0	-0	-0	-0	-0	0	-0	-0	-0	-0	-0	0.01	0.03	-0	-0	0.06	-0	1	0.06	0.08	0.02	0.03	0.01
EW45	Q18	-0	-0	0	-0.1	-0	-0	-0.1	-0	-0	0	-0.1	-0	-0	-0	-0	0.03	0.03	-0	0.12	0.07	0.06	1	0.07	0.03	0.02	-0
	Q30	-0	-0	-0	-0	-0	0.01	-0.1	-0	-0	-0	-0.1	-0	-0	-0	0	0	-0	-0	0.02	0.04	0.08	0.07	1	0.03	0.04	-0
	Q17	-0	-0	-0	-0.1	-0.1	-0	-0	-0	-0	-0	-0.1	-0.1	-0	-0.1	0.02	0.09	0.12	-0	0.03	-0	0.02	0.03	0.03	1	0.11	0.03
	Q25	-0	-0	-0	-0.1	-0.1	-0	-0	-0	-0	-0	-0.1	-0	-0	-0.1	0.01	0.03	0.08	-0	0.01	0.01	0.03	0.02	0.04	0.11	1	0.07
	Q26	-0	-0	-0	-0	-0	0	-0	0.01	-0	0	-0	-0	0.05	-0	0.02	-0	-0	-0	0.01	0.01	0.01	-0	-0	0.03	0.07	1

(a)

Transient & Expert		EW12 Item Group														EW3 Item Group				EW45 Item Group							
		Q6	Q7	Q8	Q10	Q20	Q23	Q24	Q9	Q12	Q14	Q19	Q21	Q22	Q27	Q4	Q15	Q16	Q28	Q5	Q11	Q13	Q18	Q30	Q17	Q25	Q26
	Q6	1	0.27	0.07	0.07	0.06	0.14	0.07	0.04	0.08	0.09	0.05	0.09	0.1	0.07	0.05	0	0.02	0.02	0.06	0.04	0.1	0.04	0.1	0.04	0.06	0.07
	Q7	0.27	1	0.12	0.05	0.04	0.08	0.02	0.1	0.06	0.11	0.04	0.07	0.06	0.05	0.06	0.03	0.06	0.06	0.12	0.04	0.13	0.12	0.11	0.07	0.09	0.1
	Q8	0.07	0.12	1	0.13	0.06	0.2	0.03	0.37	0.1	0.11	0.04	0.1	0.08	0.08	0.01	-0	0.02	0.04	0.1	0.04	0.09	0.1	0.09	0.06	0.07	0.08
	Q10	0.07	0.05	0.13	1	0.07	0.08	0.11	0.08	0.05	0.05	0.06	0.05	0.07	0.08	0.01	0.02	0.03	0.04	0.05	0.1	0.1	0.05	0.08	0.04	0.04	0.07
	Q20	0.06	0.04	0.06	0.07	1	0.11	0.09	0.09	0.03	0.1	0.2	0.13	0.13	0.08	0.03	-0	0.01	0.06	0.05	0.06	0.1	0.05	0.11	0.06	0.1	0.11
	Q23	0.14	0.08	0.2	0.08	0.11	1	0.4	0.15	0.09	0.16	0.11	0.26	0.3	0.16	0.02	-0	0	0.04	0.08	0.06	0.12	0.05	0.18	0.07	0.1	0.16
EW12	Q24	0.07	0.02	0.03	0.11	0.09	0.4	1	0.07	0.04	0.07	0.06	0.11	0.17	0.1	0.01	-0	0.01	0.08	0.03	0.04	0.07	0.01	0.11	0.04	0.08	0.11
	Q9	0.04	0.1	0.37	0.08	0.09	0.15	0.07	1	0.05	0.15	0.12	0.15	0.11	0.09	0.06	-0	0.03	0.07	0.13	0.06	0.13	0.1	0.12	0.09	0.12	0.14
	Q12	0.08	0.06	0.1	0.05	0.03	0.09	0.04	0.05	1	0.09	0.04	0.07	0.05	0.09	0.01	0.02	0.02	0.02	0.07	0.06	0.1	0.07	0.07	0.04	0.05	0.06
	Q14	0.09	0.11	0.11	0.05	0.1	0.16	0.07	0.15	0.09	1	0.09	0.2	0.12	0.12	0.04	-0	0.04	0.03	0.11	0.06	0.14	0.07	0.14	0.1	0.14	0.15
	Q19	0.05	0.04	0.04	0.06	0.2	0.11	0.06	0.12	0.04	0.09	1	0.12	0.12	0.09	0.04	-0	-0	0.05	0.05	0.04	0.08	0.02	0.09	0.01	0.08	0.1
	Q21	0.09	0.07	0.1	0.05	0.13	0.26	0.11	0.15	0.07	0.2	0.12	1	0.3	0.14	0.04	-0	0.02	0.07	0.12	0.07	0.11	0.07	0.15	0.06	0.12	0.15
	Q22	0.1	0.06	0.08	0.07	0.13	0.3	0.17	0.11	0.05	0.12	0.12	0.3	1	0.15	0.05	-0	0.04	0.05	0.1	0.07	0.14	0.05	0.17	0.08	0.13	0.25
	Q27	0.07	0.05	0.08	0.08	0.08	0.16	0.1	0.09	0.09	0.12	0.09	0.14	0.15	1	0.02	-0	0.04	0.07	0.06	0.03	0.09	0.03	0.12	0.03	0.11	0.12
	Q4	0.05	0.06	0.01	0.01	0.03	0.02	0.01	0.06	0.01	0.04	0.04	0.04	0.05	0.02	1	0.23	0.1	0.25	0.14	0.04	0.14	0.1	0.13	0.13	0.11	0.12
EW3	Q15	0	0.03	-0	0.02	-0	-0	-0	-0	0.02	-0	-0	-0	-0	-0	0.23	1	0.04	0.17	0.1	-0	0.09	0.05	0.05	0.07	0.03	0.02
	Q16	0.02	0.06	0.02	0.03	0.01	0	0.01	0.03	0.02	0.04	-0	0.02	0.04	0.04	0.1	0.04	1	0.09	0.11	0.08	0.13	0.12	0.12	0.23	0.23	0.16
	Q28	0.02	0.06	0.04	0.04	0.06	0.04	0.08	0.07	0.02	0.03	0.05	0.07	0.05	0.07	0.25	0.17	0.09	1	0.1	0.03	0.12	0.06	0.12	0.11	0.11	0.14
	Q5	0.06	0.12	0.1	0.05	0.05	0.08	0.03	0.13	0.07	0.11	0.05	0.12	0.1	0.06	0.14	0.1	0.11	0.1	1	0.22	0.38	0.48	0.32	0.22	0.24	0.22
	Q11	0.04	0.04	0.04	0.1	0.06	0.06	0.04	0.06	0.06	0.06	0.04	0.07	0.07	0.03	0.04	-0	0.08	0.03	0.22	1	0.27	0.22	0.24	0.13	0.16	0.15
	Q13	0.1	0.13	0.09	0.1	0.1	0.12	0.07	0.13	0.1	0.14	0.08	0.11	0.14	0.09	0.14	0.09	0.13	0.12	0.38	0.27	1	0.37	0.47	0.24	0.27	0.28
EW45	Q18	0.04	0.12	0.1	0.05	0.05	0.05	0.01	0.1	0.07	0.07	0.02	0.07	0.05	0.03	0.1	0.05	0.12	0.06	0.48	0.22	0.37	1	0.3	0.2	0.2	0.2
	Q30	0.1	0.11	0.09	0.08	0.11	0.18	0.11	0.12	0.07	0.14	0.09	0.15	0.17	0.12	0.13	0.05	0.12	0.12	0.32	0.24	0.47	0.3	1	0.22	0.27	0.26
	Q17	0.04	0.07	0.06	0.04	0.06	0.07	0.04	0.09	0.04	0.1	0.01	0.06	0.08	0.03	0.13	0.07	0.23	0.11	0.22	0.13	0.24	0.2	0.22	1	0.4	0.29
	Q25	0.06	0.09	0.07	0.04	0.1	0.1	0.08	0.12	0.05	0.14	0.08	0.12	0.13	0.11	0.11	0.03	0.23	0.11	0.24	0.16	0.27	0.2	0.27	0.4	1	0.47
	Q26	0.07	0.1	0.08	0.07	0.11	0.16	0.11	0.14	0.06	0.15	0.1	0.15	0.25	0.12	0.12	0.02	0.16	0.14	0.22	0.15	0.28	0.2	0.26	0.29	0.47	1

(b)

FIG. 9. Correlation matrices of (a) novice ($\mathbf{C}_{1/2}$) and (b) transient and expert ($\mathbf{C}_{2/2}$), made by dividing the entire ESJ samples into two, depending on the total score s^1 less than or more than -0.5 (corresponding to the raw total-score rate about $14/30$). The item number (Q) has been ordered to classify the three item groups EW12, EW3, and EW45, and the other four items of EWe are omitted. The warm collars are put on the correlation coefficients that deviate from the mean value of the corresponding correlation matrix of $\mathbf{C}_{R1}$ ($\mathbf{C}_{1/2-R1}$ or $\mathbf{C}_{2/2-R1}$) positively more than double the standard deviation of $\mathbf{C}_{1/2-R1}$ or $\mathbf{C}_{2/2-R1}$, respectively. The cool colors with lateral stripes are put on them that deviate from the corresponding mean value mentioned above negatively more than double the corresponding standard deviation, respectively.

[1] D. Hestenes, M. Wells, and G. Swackharmer, Force Concept Inventory, Phys. Teach. **30**, 141 (1992).
[2] D. Hestenes and I. Halloun, Interpreting the Force Concept Inventory, Phys. Teach. **33**, 502 (1995).
[3] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).
[4] A. A. diSessa, *Knowledge in Pieces, Constructivism in the Computer Age*, edited by G. Foreman and P. B. Putall (Erlbaum Associates, Mahwah, NJ, 1988), pp. 49–70.
[5] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).
[6] J. Minstrell, Facets of students' knowledge and relevant instruction, research in physics learning: Theoretical issues and empirical studies, in *Proceedings of the International Workshop, Bremen, Germany*, edited by R. Duit, E. Goldberg, and H. Niedderer (IPN, Kiel, Germany, 1991), pp. 110–128.
[7] F. C. Bartlett, *Remembering* (Cambridge University Press, Cambridge, England, 1932).
[8] D. E. Rumelhart, *Notes on a Schema for Stories, Representation and Understanding*, edited by D. G. Bobrow and A. M. Collins (Academic Press, New York, NY, 1975), pp. 211–236.
[9] M. S. Sabella and E. F. Redish, Knowledge activation and organization in physics problem solving, Am. J. Phys. **75**, 1017 (2010).
[10] See E. F. Redish, *Teaching Physics with the Physics Suite* (John Wiley & Sons Inc, New York, NY, 2003).
[11] D. Hammer, Misconceptions or P-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intention?, J. Learn. Sci. **5**, 97 (1996).
[12] D. Hammer, More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for educational research, Am. J. Phys. **64**, 1316 (1996).
[13] S. Graham, B. Weiner, D. C. Berliner, and R. C. Calfee, Theories and principles of motivation, in *Handbook of Educational Psychology* (MacMillan, London, 1996), pp. 63–84.
[14] D. J. Stipec, D. C. Berliner, and R. C. Calfee, Motivation and instruction, in *Handbook of Educational Psychology* (MacMillan, London, 1996), pp. 85–113.
[15] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).
[16] D. Hammer, Student resources for learning introductory physics, Am. J. Phys. **68**, S52 (2000).
[17] S. M. Stoen, M. A. McDaniel, R. F. Frey, K. M. Hynes, and M. J. Cahill, Force Concept Inventory: More than just conceptual understanding, Phys. Rev. Phys. Educ. Res. **16**, 010105 (2020).
[18] R. E. Scherr, Modeling student thinking: An example from special relativity, Am. J. Phys. **75**, 272 (2007).
[19] D. Huffman and P. Heller, What does the Force Concept Inventory actually measure?, Phys. Teach. **33**, 138 (1995).
[20] T. F. Scott, D. Schumayer, and A. R. Gray, Exploratory factor analysis of a Force Concept Inventory data set, Phys. Rev. ST Phys. Educ. Res. **8**, 020105 (2012).
[21] T. F. Scott and D. Schumayer, Conceptual coherence of non-Newtonian worldviews in Force Concept Inventory data, Phys. Rev. Phys. Educ. Res. **13**, 010126 (2017).
[22] M. R. Semak, R. D. Dietz, R. H. Pearson, and C. W. Willis, Examining evolving performance on the Force Concept Inventory using factor analysis, Phys. Rev. Phys. Educ. Res. **13**, 010103 (2017).
[23] P. Eaton and S. D. Willoughby, Confirmatory factor analysis applied to the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. **14**, 010124 (2018).
[24] C. Distefano, M. Zhu, and D. Mindrila, Understanding and using factor scores: Considerations for the applied researcher, Pract. Assess. Res. Eval. **14**, 20 (2009).
[25] F. M. Lord, *Application of Item Response Theory to Practical Testing Problems* (Lawrence Erlbaum Associates, Mahwah, NJ, 1980).
[26] M. Planinic, L. Ivanjek, and A. Susac, Rash model-based analysis of the Force Concept Inventory, Phys. Rev. ST Phys. Educ. Res. **6**, 010103 (2010).
[27] J. Wang and L. Bao, Analyzing Force Concept Inventory with item response theory, Am. J. Phys. **78**, 1064 (2010).
[28] R. D. Bock, Estimating item parameters and latent ability when responses are scored in two or more nominal categories, Psychometrika **37**, 29 (1972).
[29] T. I. Smith, K. J. Louis, B. J. Ricci, and N. Bendjilali, Qualitatively ranking incorrect responses to multiple-choice questions using item response theory, Phys. Rev. Phys. Educ. Res. **16**, 010107 (2020).
[30] T. F. Scott and D. Schumayer, Students' proficiency scores within multi-trait item response theory, Phys. Rev. ST Phys. Educ. Res. **11**, 020134 (2015).
[31] J. Stewart, C. Zabriskie, S. DeVore, and G. Stewart, Multidimensional item response theory and the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. **14**, 010137 (2018).
[32] P. Eaton and S. Willoughby, Identifying a preinstruction to postinstruction factor model for the Force Concept Inventory within a multitrait item response theory framework, Phys. Rev. Phys. Educ. Res. **16**, 010106 (2020).
[33] J. Stewart, B. Drury, J. Wells, A. Adair, R. Henderson, Y. Ma, A. Perez-Lemonche, and D. Pritchard, Examining the relation of correct knowledge and misconceptions using the nominal response model, Phys. Rev. Phys. Educ. Res. **17**, 010122 (2021).
[34] G. A. Morris, L. B.-Martin, S. D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley, Testing the test: Item response curves and test quality, Am. J. Phys. **74**, 449 (2006).
[35] G. A. Morris, N. Harshman, L. B.-Martin, E. Mazur, T. Mzoughi, and S. D. Baker, An item response curves analysis of the Force Concept Inventory, Am. J. Phys. **80**, 825 (2012).
[36] M. Ishimoto, H. Nitta, and R. Lang, 2014 survey on physics education in Japan (2): Item response curves of the FCI data, in *Proceedings of the Meeting Report of the 2015 Annual Meeting of the Physical Society of Japan, 23 pCK-3 (Written in Japanese)* (2015), 10.11316/jpsgaiyo.70.1.0_3482.
[37] T. Yamazaki, S. Kishizawa, Y. Hasegawa, J. Yasuda, H. Minoda, M. Goda, T. Murata, and R. Lang, Analysis of

mechanics conceptual understanding through "item response curves" I, in *Proceedings of the Meeting Report of the 2017 Annual Meeting of the Physical Society of Japan, 19 pC11-3 (Written in Japanese)* (2017), 10.11316/jpsgaiyo.72.1.0_3376.

[38] S. Kishizawa, T. Yamazaki, Y. Hasegawa, J. Yasuda, H. Minoda, M. Goda, T. Murata, and R. Lang, Analysis of mechanics conceptual understanding through "item response curves" II, in *Proceedings of the Meeting Report of the 2017 Annual Meeting of the Physical Society of Japan, 19 pC11-3 (Written in Japanese)* (2017), 10.11316/jpsgaiyo.72.1.0_3377.

[39] M. Ishimoto, G. Davenport, and M. C. Wittman, Use of item response curves of the force and motion conceptual evaluation to compare Japanese and American students' views on force and motion, Phys. Rev. Phys. Educ. Res. **13**, 020135 (2017).

[40] J. C. Gower and D. J. Hand, *Biplots* (Chapman and Hall, London, 1995).

[41] G. H. Dunteman, *Principal Component Analysis* (SAGE, Beverly Hills, 1989).

[42] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics (Springer, Berlin, 2002).

[43] J. C. Heyton, D. G. Allen, and V. Scarpello, Factor retention decision in exploratory factor analysis: A tutorial on parallel analysis, Organ. Res. Methods **7**, 191 (2004).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/xw7r-2rx1 for the data on the artificial inter-item-uncorrelated samples of R1.

[45] D. J. Law and K. F. Wilson, The role of competing knowledge structures in undermining learning: Newton's second and third laws, Am. J. Phys. **85**, 54 (2017).

[46] K. Mizuno, *The Principle of Multivariate Data Analysis (Written in Japanese)* (Asakura Shoten, Tokyo, 1996).

[47] M. Goda, S. Kishizawa, and Y. Hasegawa, Revisit the FCI results with PCA I (PCA and FA), in *Proceedings of the Meeting Report of the 2019 Semi-Annual Meeting of the Physical Society of Japan 11pB24-8 (Written in Japanese)* (2019), 10.11316/jpsgaiyo.74.2.0_2816.

[48] M. Goda, S. Kishizawa, and Y. Hasegawa, Revisit the FCI results with PCA II (Principal component score and the divisions), in *Proceedings of the Meeting Report of the 2019 Semi-Annual Meeting of the Physical Society of Japan 11pB24-9 (Written in Japanese)* (2019), 10.11316/jpsgaiyo.74.2.0_2817.

[49] M. Goda, S. Kishizawa, and Y. Hasegawa, Revisit the FCI results with PCA III (PC expansion of the correlation between item scores), in *Proceedings of the Meeting report of the 2020 Annual Meeting of the Physical Society of Japan 19 pK28-4 (Written in Japanese)* (2020), 10.11316/jpsgaiyo.75.1.0_3010.

[50] M. Goda, S. Kishizawa, and Y. Hasegawa, Inter-item correlation of the correct answers of FCI, in *Proceedings of the Meeting Report of the 2021 Annual Meeting of the Physics Education Society of Japan (Written in Japanese)* (2021), pp. 50–51, https//pesj.jp/2021doc/Yokou2021.pdf.

[51] S. Kanim and X. C. Cid, Demographics of physics education research, Phys. Rev. Phys. Educ. Res. **16**, 020106 (2020).

Correction: A misprint in the section heading for II B introduced during the production cycle has been fixed.