Conceptual Coherence in Force Concept Inventory Data

T F Scott and D Schumayer
Department of Physics, University of Otago, Dunedin 9106, New Zealand
terry.scott@otago.ac.nz

Abstract. The Force Concept Inventory (FCI) has been used as a research instrument and to evaluate the conceptual understanding of students since it was first published in 1992. It is commonly used to assess the formation of conceptual structures in the minds of students who are learning Newtonian dynamics. For this reason it is vital that both researchers and teachers are assured that such conceptual coherence does, in fact, appear in FCI data, and are aware of the differences between Newtonian concepts as they appear in experts, and as they appear in students. The research presented here provides evidence that this conceptual coherence exists in FCI response data. This evidence is the result of a careful factor analysis of such data and we present the factor structure found in this analysis. This factor structure does not correspond exactly to the conceptual structure proposed by the authors of the FCI and thus is evidence of the difference between expert and novice conceptualisations of Newtonian Mechanics. Furthermore, we also provide an item response analysis of FCI data which is able to provide us with a better understanding of the specific abilities developed by our students and the interactions between these abilities as student obtain mastery of Newtonian ideas.

1. Introduction
The Force Concept Inventory was introduced in 1992 by Hestenes, Wells, and Swackhamer [1] and was updated in 1995 [2]. The primary use of the FCI has been to evaluate the conceptual understanding of Newtonian dynamics among students and to give a quantitative measure of the effectiveness of teaching methods in this area. The FCI claims to measure the concepts possessed by students, however this claim was debated soon after the publication of the original paper [3]. Huffman and Heller [3] performed a factor analysis on FCI data collected from two groups, a group of 145 high school students and a group of 750 university students. This analysis found no evidence of underlying conceptual coherence in this FCI data. In other words, their analysis found as many independent factors in the response data as there are questions in the survey.

The research presented here addresses the issue of conceptual coherence in FCI response data. Our analysis will confirm the appearance of coherence in this data and we will further analyse this coherence to provide some insights into the conceptual formations possessed by students who are in the process of learning Newtonian dynamics. To achieve this we will present our own factor analysis and a multi-trait item response analysis of the same FCI data. The research in this talk was originally published in two papers [5,6].

2. Collection of Data
FCI data was collected from students enrolled in a one semester physics service course. The data was collected over two years (2008 and 2009). This is a traditional, algebra based course taken by students intending to enter a variety of health science professional programs (e.g., medicine, dentistry,
Students were given the opportunity to complete the FCI at the end of the mechanics section of the course, but completing the FCI was not mandatory nor was the grade included in their internal assessment. The FCI was also completed online in the students’ own time.

This data collection method is problematic in some respects, namely there is no direct control over the time students take to complete the survey and the motivation to answer questions correctly is not due to internal assessment rewards. These problems are mitigated by the fact that this course is a component of a competitive entry programme and as such the majority of the students are already significantly motivated to do well in the survey. Prior to any statistical analysis the data was checked and frivolous attempts removed.

Over the two years in which data was collected (2008 and 2009) a total of 2150 records were collected. The overall score ranged from 0 to 30, the mean score was 15.04 (median 14), and the standard deviation was 6.16. Only 41 (< 2%) student responses were excluded as being frivolous.

3. **Factor Analysis**

Factor analysis is well known and commonly used technique in the statistical analysis of psychometric and educational data. There are a number of discussions of factor analysis in the research literature and we will only indicate the main ideas underlying the technique here. A good textbook introduction to factor analysis is given by Kim and Meuller [6]. The value of factor analysis to our project is that it allows us to identify coherence in FCI data. This coherence may be attributed to conceptual formations in the minds of students, but it is generally safer to use the neutral term “latent trait” for the underlying cause of the coherences seen in response data.

The factor analysis reported here is exploratory, this means that the factors are determined by the data. To perform this analysis, a correlation matrix between the questions of the FCI is constructed. Factor analysis amounts to finding the eigenmatrices and eigenvalues of this matrix. The questions which appear in an eigenmatrix are grouped together in a factor, and the eigenvalue of this matrix is the factor loading for these questions. We have factored the correlation matrix into non-orthogonal matrices.

This means that the correlation function chosen for the data is of critical importance. The correlation function appropriate to binary categorical data is the tetrachoric correlation function. The calculation of the tetrachoric correlation function is quite complex and involves a number of assumptions regarding the nature of the data under analysis. The data in our analysis satisfies these assumptions, further information about the use of the tetrachoric correlation function in factor analysis may be found in the discussion provided by Schmitt [7] and for more technical details, see Refs. [8–11].

4. **Factor Analysis - Conclusions**

The primary result of this analysis is that FCI data does indeed display coherence, the evidence for this is the appearance of stable, interpretable factors. Our analysis indicates that that FCI response data supports the use of either a single-factor model, or a five-factor model. In both cases, the factor models account for about 40% of the variation in the data set. The five factor model produces interpretable factors, which are listed below.

| **Factor Analysis**   | **Hestenes et al.** |
|-----------------------|---------------------|
| 1. Identification of forces. | 1. Kinematics |
| 2. Newton’s first law with zero force. | 2. First Law |
| 3. Newton’s second law and kinematics. | 3. Second Law |
| 4. Newton’s first law with cancelling forces. | 4. Third Law |
| 5. Newton’s third law. | 5. Superposition Principle |
|                       | 6. Kinds of Force |
We also include the division of the Newtonian force concept into “dimensions” supplied by the authors of the original paper. It is important to note that our factors represent the coherences in the data presented by students, this does not invalidate the division into dimensions provided by Hestenes et al. [1], as this represents the understanding of experts rather than novices.

Our factor analysis indicates the presence of significant correlations among the factors. Factor 1 is strongly correlated with Factors 2, 3, and 4. Factor 5 does not correlate strongly with any of the other factors. These correlations are understandable in that Factor 1, which is associated with the identification of forces, is needed before the other factors can engage. Factor 5, which is associated with the third law, is largely independent of the other factors in the minds of students.

It appears that Newton’s third law is closely, but incorrectly, associated with Newton’s first law. This is shown by the appearance of question 16 (a third law question) in the factors associated with the first law. This appears to be related to the fact that question 16 relates to a constant velocity situation and students will assume that forces sum to zero because of Newton’s first law, rather than because they are action-reaction pairs. This suggests that great pains should be taken to distinguish problems requiring Newton’s third law from problems which require Newton’s first law.

Finally, it also appears that the ideas of kinematics are not well differentiated by students from Newton’s first and second laws. In particular, students seem to struggle to clearly identify kinematic ideas in the areas of dynamics in which the former concepts are used. This suggests that kinematic concepts should be clearly distinguished from dynamical concepts during the introduction of Newton’s first and second laws.

5. Item Response Analysis

Item response analysis is another latent trait method which is well attested in the literature [12-15], and there are many excellent technical monographs explaining this technique (see for example [16]). It is a latent trait method in that it assumes that there are hidden cognitive properties, possessed by students, which result in the response data collected by surveys such as the FCI. In the case of latent trait analysis it is postulated that these traits effect the probability of a correct response by a student. The probability of a correct response is given by the formula:

\[
P(x_{ij} = 1|\theta_i, \{a_j, d_j, g_j\}) = g_j + \frac{1 - g_j}{1 + \exp(-a_j\theta_i - d_j)}
\]

In this formula, \(P\) is the probability that the \(i\)-th student will answer the \(j\)-th question correctly. \(\theta_i\) is the trait level possessed by the \(i\)-th student (we will call this the student’s proficiency), \(g_j\) is the “guessing” parameter of the \(j\)-th question (it’s the probability that a student will guess the correct answer), \(a_j\) is the “discrimination” parameter of the \(j\)-th question, and \(d_j\) is the “difficulty” parameter of the \(j\)-th question. The item response analysis involves fitting these parameters to the response data. Parameters are chosen to optimise the match between the scores predicted by the model and the actual scores recorded in the data. These parameters are illustrated in the diagram below.

The value of item response analysis is that it characterises the questions in the survey but also provides an optimised measure of student ability. The proficiency of the student gives the position of the student on the horizontal axis of the item response diagram (see Figure 1 below). The proficiency is roughly the degree to which the student instantiates the relevant underlying trait. Proficiency is calibrated against the item parameters so that a proficiency equal to \(-d/a\) will mean that that student has a 50% chance of responding correctly to that question.
This formula and diagram represent the probability of a correct response by students when the model contains a single underlying latent trait. However, our earlier analysis shows that FCI data may profitably be modelled using five underlying traits. For this reason our item response analysis is expanded to include the possibility of multiple underlying latent traits. In these models the probability of a correct response by the $i$-th student to the $j$-th question is given by the formula,

$$P(x_{ij} = 1 | \theta_i, \{a_j, d_j, g_j\}) = g_j + \frac{1 - g_j}{1 + \exp(-a_j \cdot \theta_i^T - d_j)}$$

In the multi-trait model there is a single guessing parameter, $g_j$, and a single difficulty parameter, $d_j$. But now the trait level, $\theta_i$, and the discrimination parameter, $a_j$, are vectors, the components of which are the values for each trait corresponding the $i$-th student and the $j$-th question respectively.

Multi-trait item response analysis requires that an exploratory factor analysis be carried out as a preliminary step to fitting item response curves to the data. This factor analysis is carried out using an alternative methodology to the factor analysis which we carried out earlier. Thus the multi-trait analysis serves as a useful check of our earlier factor analysis. The latent traits identified by the multi-trait item response analysis correspond closely to the factors we identified in our earlier exploratory factor analysis. For the sake of later convenience we will change the numbering of the factors found in the multi-trait item response analysis compared to the numbering used in the factor analysis. To avoid confusion we will refer to the factors in the item response analysis as traits. The traits referred to in the item response analysis are

- Trait 1: Newton’s first law with zero force.
- Trait 2: Newton’s third law.
- Trait 3: Identification of forces.
- Trait 4: Newton’s first law with cancelling forces.
- Trait 5: Newton’s second law and kinematics.

![Figure 1. The single trait item response curve [5].](image-url)
6. Multi-trait Item Response Analysis - Conclusions

As mentioned previously, one distinctive feature of item response analysis is the assignment of proficiency score (or latent trait levels) to students. This gives the researcher an additional tool in the analysis of FCI data. One is able to list the proficiencies of students in each of the underlying traits included in the analysis and further, it is relatively straightforward to then investigate correlations between these proficiencies. In the study presented here the correlations between the proficiencies were calculated and are shown in the following table.

| Traits from Item Response Theory | 1  | 2  | 3  | 4  | 5  |
|---------------------------------|----|----|----|----|----|
| 1                               | 1.000 |
| 2                               | 0.519 | 1.000 |
| 3                               | 0.769 | 0.643 | 1.000 |
| 4                               | -0.738 | -0.602 | -0.756 | 1.000 |
| 5                               | -0.719 | -0.624 | -0.770 | 0.704 | 1.000 |

The significant result of this analysis is the clear separation of the trait proficiencies into two distinct groups. Proficiency in traits 1 to 3 are quite strongly correlated with each other, similarly the proficiency in traits 4 and 5 are strongly correlated with each other. However, the correlation between these two groups of proficiencies is equally strong but negative, e.g., the correlation between proficiency in trait 1 and 3 is 0.769, but the correlation between proficiencies in trait 1 and 5 is -0.719.

The first group includes traits 1, 2, and 3. These are proficiency at identifying forces, proficiency at applying Newton’s first law with zero force, and proficiency at applying Newton’s third law. It is not surprising that identification of forces and Newton’s third law are in the same group, since an understanding of the third law requires an added level of sophistication in the identification of forces. However, it is somewhat surprising that Newton’s first law with zero force is also in this group. It is possible that this is related to the effect observed with question 16. In the factor analysis it was found that question 16 appeared in the first law factors, even though it is clearly a questions concerning Newton’s third law. It would appear that students are erroneously applying Newton’s first law to this situation and consistently arriving at the correct answer. Thus it may well be that mastery of Newton’s first law with zero force facilitates an understanding of Newton’s third law.

The second grouping contains proficiency in traits 4 and 5. These traits contain Newton’s first law with cancelling forces (trait 4) and Newton’s second law and kinematics (trait 5). It is again unclear why proficiencies in these two traits should be strongly associated.

It is also surprising that there is a strong anti-correlation between these two groups of proficiency and, clearly, further research in this area would be fruitful. One possible explanation is suggested by the observation that the questions which appear in traits 1, 2 and 3 appear to emphasise visual reasoning. Traits 4 and 5, on the other hand, contain questions which appear to be more “word heavy”. It is possible that these two groupings reflect different thinking styles. This would explain why the two first law traits are in different groups and proficiencies in these traits are strongly anti-correlated. Again, this can be no more than speculation at this point and more research is required.

7. References

[1] Hestenes D et al. 1992 Phys. Teach. 30 141 https://doi.org/10.1119/1.2343497
[2] Available from http://modeling.asu.edu/R&E/Research.html
[3] Huffman D and Heller P 1995 Phys. Teach. 33 138 https://doi.org/10.1119/1.2344171
[4] Scott T F et al. 2012 Phys.Rev. ST Phys. Educ. Res. 8 020105 https://doi.org/10.1103/PhysRevSTPER.8.020105
[5] Scott T F and Schumayer D 2015 *Phys. Rev. ST Phys. Educ. Res.* 11 020134 https://doi.org/10.1103/PhysRevSTPER.11.020134

[6] Kim J-O and Meuller C W 1978 *Introduction to Factor Analysis: What it is and How to do it* (Beverly Hills, California: SAGE).

[7] Schmitt T A 2011 *J. Psychoeduc. Assess.* 29 304 https://doi.org/10.1177/0734282911406653

[8] Hamdan M A 1970 *Biometrika* 57 212 https://doi.org/10.1093/biomet/57.1.212

[9] Digby P G N 1983 *Biometrics* 39 753 https://doi.org/10.2307/2531104

[10] Mislevy R J 1986 *J. Educ. Stat.* 11 1 https://doi.org/10.3102/10769986011001003

[11] Bonett D G and Price R M 2005 *J. Educ. Behav. Stat.* 30 213 https://doi.org/10.3102/1076998603002213

[12] Planinic M et al. 2010 *Phys. Rev. ST Phys. Educ. Res.* 6 010103. https://doi.org/10.1103/PhysRevSTPER.6.010103

[13] Wallace C S and Bailey J M 2010 *Astron. Educ. Rev.* 9 010116 https://doi.org/10.3847/AER2010024

[14] Planinic M et al. 2006 *J. Res. Sci. Teach.* 43 150 https://doi.org/10.1002/tea.20101

[15] Han J et al. 2015 *Phys. Rev. ST Phys. Educ. Res.* 11 010112 https://doi.org/10.1103/PhysRevSTPER.11.010112

[16] Baker F B and Kim S-H 2004 *Item Response Theory: Parameter Estimation Techniques* (New York: Marcel Dekker).