Analysis of Prerequisites: Methodology and a Case Study

Mike Lopez, Dobrila Lopez, Chris McCarthy, Robert Oliver

Christchurch Polytechnic Institute of Technology
130 Madras Street,
Christchurch, New Zealand
+64 3 940 8000

lopezm@cpit.ac.nz, lopezd@cpit.ac.nz, mccarthyc@cpit.ac.nz, oliverr@cpit.ac.nz

ABSTRACT

It is well known that it is dangerous to infer causation from correlation. However, the mantra that correlation does not imply causation can lead to some researchers believing that formal inference is never possible from a correlational study. This paper presents a theoretical framework, a conceptual framework and a methodology for establishing formal inference from the analysis of prerequisites in an educational context. This is important in education because some prior knowledge is often required for success in any topic or course. The method is illustrated with a case study that investigates the effectiveness of a level four certificate as preparation for further study. The case study identified the unique contribution to subsequent performance made by individual courses in the certificate. It also identified the specific courses in subsequent study which were most affected by the certificate courses. We conclude that the approach can indeed enable formal inference from a correlational study.

Categories and Subject Descriptors

K.3.2 [Computer and Information Science Education]

General Terms

Measurement

Keywords

Correlation and causation, causal inference, prerequisite analysis.

1. INTRODUCTION

At an early stage in their career, researchers encounter the mantra: *correlation does not imply causation*. Indeed, naively attributing causation is subject to at least two logical fallacies. The first, known as *cum hoc ergo propter hoc*, (with this, therefore because of this), arises when the assumption is made that when two events co-occur, one must cause the other. The second, known as *post hoc ergo propter hoc* (after this, therefore because of this) arises when the assumption is made that an event that follows another is necessarily a consequence of the first event.

Given these potential logical fallacies, several schools of thought arise. The first is that causation can only be asserted when there is an experimental study design, which is carried out under controlled conditions, and in which the causative variable is manipulated experimentally. There is no doubt that this can give good evidence of causation. However, it also precludes studying causation in such diverse scientific fields as meteorology, oceanography, cosmology or neuroscience, in which experimental manipulation under controlled conditions is not possible. These have to be studied "in the wild", as it were. The second school of thought is concerned with investigating causation in this context. The essence of this school of thought is the use of a theory driven approach. A theory is constructed which includes causative mechanisms and explains the observations. This theory is then put to the test. Theories that survive testing are regarded as plausible until disconfirmed and they gain credibility as further tests are carried out [1]. Examples of this paradigm are Hawking radiation from black holes, predicted in 1974 [2] and partially confirmed in 2012 [3] or the search for the Higgs Boson, predicted in 1964 [4] and tentatively confirmed in 2012 [5]. Few would disagree that these researches were scientific endeavours.

The third school of thought arises when researchers, aware of the difficulties of causal inference, claim to be studying associations rather than causation. Inevitably, though, we are usually interested in explanation and causation rather than just associations. As Michael Rutter notes:

> Researchers tend to fall into one of two camps with respect to how they react to the problem. First, there are those who are careful to use language that avoids any direct claim for causation, and yet, in the discussion section of their papers, they imply that the findings do indeed mean causation. Second, there are those that completely accept the inability to make a causal inference on the basis of simple correlation or association and, instead, take refuge in the claim that they are studying only associations and not causation. This second, ''pure'' approach sounds safer, but it is disingenuous because it is difficult to see why anyone would be interested in statistical associations or correlations if the findings were not in some way relevant to an understanding of causative mechanisms. [19, p. 377]

In education, ethical considerations usually preclude the use of experimental study designs. We have a duty of care to our students and therefore cannot deliberately expose students to experimental conditions that we believe might be less than optimal for their learning. Moreover, ethical approval for any research is likely to involve the fundamental principles of informed consent and no deception. It is difficult to conceive of any meaningful educational intervention or treatment that could be studied in a way that preserves the double-blind nature essential to the validity of experiments. This is essential to control common threats to validity such as the Hawthorne effect [6]. It follows that correlational studies are a natural choice for educational research. In a sense, we have to carry out our studies "in the wild" even though this brings in many possible confounding factors. Despite the problems, there are however, some benefits of this approach. For example, such studies have greater ecological validity and they keep research findings close to practice.

This quality assured paper appeared at the 4th annual conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2013) incorporating the *26h Annual Conference of the National Advisory Committee on Computing Qualifications*, Hamilton, New Zealand, October 6-9, 2013. Mike Lopez and Michael Verhaart, (Eds).

In designing such a study, assessment data are a natural source of objective evidence. Using these data avoids the additional burden placed on students by gathering extra research data. Moreover, these data are often the best quality data available to educators because most institutions invest substantial resources in the quality control of these data, typically using formal moderation procedures both before and after administration of the assessment. Moreover, as Timperley and Parr [7] note, educators:

> … spend a great deal of time assessing students more formally and recording and reporting the results, but research seems to reflect that this information is not necessarily used to inform teaching decisions. [22, p.12]

From this perspective, it is important to make use of such data, wherever possible, to inform our teaching. However, given the potential logical fallacies, it is clearly important to tread carefully when building inference from a correlational study. Figure 1 illustrates one of the difficulties with subtle humour.

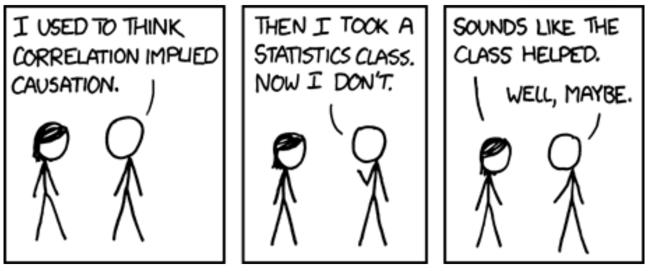

Figure 1: Correlation and causation. [1]

Of course, it is easy and entertaining to poke fun at such inference and to create contrived examples to point out the logical fallacies. Figure 2 shows a plot of the global average temperature against an approximate estimate of the number of pirates in the world.

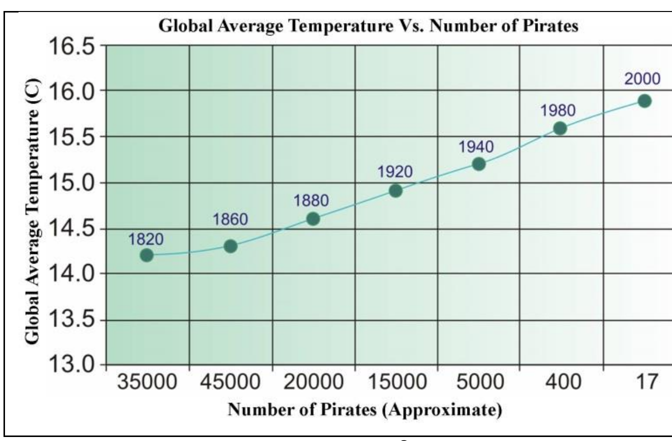

Figure 2: pirates and global warming[2]

One tongue-in-cheek explanation of the association is that global warming evaporates rum and this, in turn, leads to a decline in the number of pirates.

Putting aside humour, we can state the general inference problem As follows. If, for two variables A and B, A correlates with B, then:

- A may cause B,
- B may cause A,
- A and B may both be caused by a common variable C, or
- There may be a complex web of causation [8] which includes both A and B.

Choosing plausible options from such possible explanations of causation is often complex and potentially challenging, but need not be impossible. For example, if event B always follows event A in time, as is typically the case with a prerequisite relationship, then it is reasonable to exclude the possibility that the later event B is a cause of A. In practice, however, such logical analysis can only take us so far and it is usually difficult to distinguish between the cases where A is a cause of B, and where both A and B are the consequence of some other, possibly unknown, factor C. Thus, moving our knowledge forward requires the development and testing of theory that can explain the causal mechanisms underpinning observed correlations.

Our method follows the second school of thought we have set out: that of theory driven exploration. In education, we often arrange learning opportunities in a way that successive material builds on what has been learned before. Typically, we do this in two ways. First, within a course, we organise and sequence material in a way that builds skills and knowledge progressively. Second, a course may assume skills and knowledge from a previous course, often as a formal prerequisite. In essence then, our tacit theory is that knowledge, both knowing what and knowing how, can be built progressively on prior understandings. This is nothing new. Indeed, the very use of the term *prerequisite* implies such a theory underpinning our practice. The important point is that we can make such a theory explicit and derive testable propositions that arise from the theory. This allows us to put such a theory to the test. For this paper, we will call our theory: the *theory of progressive knowledge building*.

Within a course, we usually have multiple assessments covering multiple topics and multiple items for each topic. It is therefore possible to derive from this a topic by topic performance analysis. Between courses the final outcome grades serve a similar purpose. These data allow us to test our theory of progressive knowledge building, either disconfirming it, or raising our confidence in its generality [1]. The essential point of this approach is that statistical methods do not, and cannot, explain phenomena. Theories provide explanations and statistical methods put theories to the test.

In summary, the idea of prerequisites is pervasive in education. We organise and present our learning experiences in a way that successive material progressively builds on what has been learned before. Our research question is: *How can we evaluate the effectiveness of such a prerequisite?* Because ethical reasons preclude experimental approaches, we are left with observational approaches. This leads us to ask how we can answer such a question from a correlational study. Our approach is to posit a theory of progressive knowledge building and then put this theory to the test.

The remainder of this paper is organised as follows. We present the theoretical framework for the analysis in section 2 and our conceptual model in section 3. We describe our method in section 4 and then illustrate the method by presenting a case study in section 5. Finally, in section 6, we present our conclusions.

---

[1] Reproduced under the terms of the Creative Commons Attribution-NonCommercial 2.5 License, from http://xkcd.com/552/

[2] Reproduced under the terms of the GNU Free Documentation License, Version 1.2 from: http://en.wikipedia.org/wiki/File%3aPiratesVsTemp_English.jpg

2. THEORETICAL FRAMEWORK

Although, following Stevens [9, p. 677], many researchers simply assign scores as numerals and assume that they are linear measurements, nevertheless at best scores only form an ordinal scale [10]. However, ordinal scores may be converted to linear measurements under mild and verifiable assumptions [11, 12, 13], or simply treated as linear to isolate a main effect where this is believed to provide sufficient information compared to a full model [14]. However, merely assigning numbers does not make an attribute quantifiable [15]. A formal framework for the scientific task of establishing that an attribute is quantifiable is given by Luce and Tukey [16], and tests for this are given by Karabatsos [17] and Lopez [12].

Taken together, the foregoing work allows linear measurement to be constructed or assumed, and all assumptions verified. Once linear measurement has been established, analysis can proceed under the general linear model. A standard approach is to use multiple-regression to investigate the extent to which a set of predictors can explain variability of a criterion variable. However, real datasets may be heavily confounded. Where a confounding variable can be identified and measured, its effect can be statistically controlled by treating it as a covariate and using the partial correlation [18]. However, when there are correlations among the predictors, a more appropriate control is to use the semi-partial correlation [19].

An approach using multiple-regression can thus identify the overall extent to which a set of predictors explains the variability of a criterion variable, together with a decomposition that estimates the unique contribution of each predictor, and any commonality shared among the predictors.

3. CONCEPTUAL MODEL

In education, many factors contribute to success in a course. We can identify three broad sets of factors that contribute to performance. The first set relates to the nature of the topic itself, including the skills and knowledge needed. The second relates to the teacher: how the material and learning is organised and how the teacher engages with the class. The third comprises the personal attributes of the student, such as the degree of effort they make to learn, their approach to studying, self-management, health, and many other elements.

In the context of studying prerequisites, we can group these factors into those shared across topics and those unique to topics. Figure 3 summarises this conceptual model.

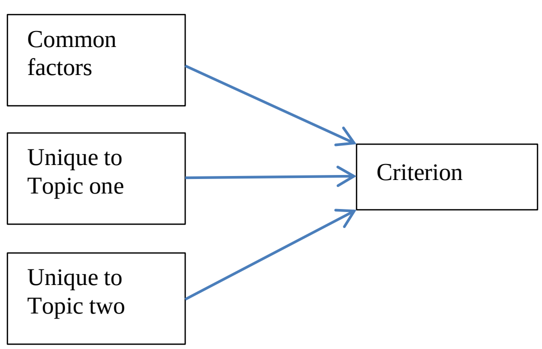

Figure 3: Conceptual model

One effect of using multiple-regression is that any linear transformation of criterion or predictor variables leads to the same result. This gives a natural statistical control of several aspects of the data; variability of performance is analysed rather than any absolute level. We can expect students' personal attributes to contribute to performance in all the predictors as well as to performance in the criterion. There is therefore a natural statistical control of these attributes. Similarly, there is an automatic statistical control of aspects such as the leniency or severity of marking in a course. Where the events measured by the criterion variable occur at a different time from those of the predictors, there may also be a time-related effect. In particular, the criterion events usually occur later and so, the student will have engaged in more learning (or have forgotten material) by this time. However, there is, again, a natural statistical control of this.

The skills and knowledge required in a topic can be decomposed into two sets: those unique to the topic and those shared across topics. Similarly, the effect of the teacher can be decomposed into effects unique to the topic and effects shared across topics.

In summary, the above points give the following decomposition. The commonality comprises skills and knowledge shared across the topics, together with any teacher characteristics shared among the teaching team. The unique topic contribution comprises the unique skills and knowledge employed in the topic, together with any unique contribution of the teacher to the topic. Student attributes, marking leniency or severity, and elapsed time effects are automatically controlled statistically.

It should be noted that for both the unique contributions and the commonality, teacher characteristics are confounded with the skills and knowledge intrinsic to a topic. One obvious refinement would be to decompose these so as to isolate the unique effects of the teacher from skills and knowledge. However, we believe that this would not be useful, and might indeed be misleading, for the following reasons. First, how the teacher engages with the class is likely to depend as much on how the teacher relates to the topic as on intrinsic teacher attributes. Isolating the latter would still leave a substantial component which is confounded with intrinsic skills and knowledge. Second, attempting to isolate intrinsic teacher characteristics would require sampling a larger set of courses with which the teacher engaged, thus limiting the applicability of the method. Third, the analysis could be readily misinterpreted as controlling for teacher effects, whereas there would still a substantial component that remains confounded. With these points in mind, and particularly with regard to the possibility of misinterpretation, we do not recommend attempting to isolate teacher effects.

4. METHOD

Although the underlying logic is the same, we present below separately the logic for investigating the relationships among topics within a course and the relationship between courses for which a formal prerequisite relationship exists or an informal relationship is assumed.

To investigate topics in a course, we assume that the course has several assessments, that each assessment covers a range of topics, and that at least some topics are assessed in more than one assessment. It is then possible, for at least some topics, to estimate performance in the topic on at least two occasions by using separate scales formed for each occasion. If the theory of progressive knowledge building is correct, there should be a stronger association between occasions for a topic, than between different topics and this can be readily tested.

For prerequisite relationships between courses, the final outcome grades can be used. It is assumed that the analysis includes both courses with a known or expected strong prerequisite relationship and others with for which the relationship is weaker or expected to be absent. This will enable the separation of performance intrinsic to a specific prerequisite course from commonalities and thus enable testing of the theory of progressive knowledge development.

In both cases, there is likely to be significant confounding of any cause and, accordingly, careful analysis is needed to unpick the causal mechanisms and interpret any findings.

5. CASE STUDY

To illustrate the method, we present here a case study of the analysis of courses in a certificate programme which also serves as a preparatory course for further study in degree or diploma courses. For this purpose, a course is equivalent to a topic as discussed above. Our working research question was: *How well does the CICT prepare students for further study?*

5.1 Courses

The Certificate in Information and Communication Technologies (CICT) consists of four courses at level four in the NZQA framework [20]. Some students complete their CICT and then leave; others progress to further study. We were interested in those students who continued their study in the Bachelor of Information and Communication Technologies (BICT) or in the Diploma of Information and Communication Technologies (Dip/ICT). Both of these programmes share a common set of papers in their first year. We focused on those courses which are taken by a student in their first semester of study in these later programmes. The courses we used are shown in Table 1.

Table 1: Courses

Code	Name
CICT400[a]	Introduction to IT
CICT410[a]	Practical skills in IT
CICT420[a]	Information and Communication Skills
CICT430[a]	Problem Solving in ICT
CS153[b]	Computer Architecture
IT181[b]	Information Technology Concepts & Tools
PR109[b]	Programming Precepts
SE101[b]	Software Engineering 1A

Note: a) CICT b) BICT or Dip/ICT

5.2 Sample and Data

Our source data were taken from our institution's student management system. We included only those CICT students who subsequently chose the further study in the degree or diploma, and only those diploma and degree students who had previously completed the CICT. We selected three cohorts of students as shown in Table 2. These were the most recent cohorts available at the time of writing this paper.

Table 2: Source data

	Cohort 1	Cohort 2	Cohort 3
CICT	2011 Sem 1	2011 Sem 2	2012 Sem1
Further study	2011 Sem 2	2012 Sem1	2012 Sem2

For each cohort of students, we collected their grades in the CICT programme and their grades in the first four papers of their subsequent study in the degree or diploma. The data used were the mark, student identifier, course and cohort.

5.3 Analysis

For screening purposes, all data were processed using a formal measurement model [12] for two scales: CICT and further study. This screening allowed all of the measurement assumptions to be tested formally. Both scales met all measurement assumptions with the exception of local dependence and differential item functioning (DIF) [21] by cohort. Neither is critical because the measurement software can detect and correct for these. Nevertheless, we note that the detected DIF indicates that the criteria for CICT400, CICT430 and CS153 changed significantly across the cohorts. Local dependence was detected among all the CICT courses and between CS153 and IT181. Moreover, local dependence was detected among seven students in CICT. Discussion of the impact of local dependence is outside the scope of this paper. Interested readers can get more information in Lopez [12]. Nevertheless, such diagnostics are illustrative of the extensive information provided to researchers by the use of a formal measurement model.

Overall, we deemed the scales suitable for analysis. The reliability of the scales is shown in Table 3.

Table 3: Reliability of scales

Scale	Alpha	PSR	ISR
CICT	.716	.817	.980
Further Study	.854	.874	.971

Note: Alpha is Cronbach's alpha, PSR is the Rasch person separation reliability, ISR is the Rasch item separation reliability.

In statistics, reliability is defined as the variance of the true score (T) expressed as a proportion of observed variance (X). This can be stated as a formula: $r = T/X$. Essentially, it expresses the accuracy of measurement. Using an analogy from electrical engineering, a reliability of .9 represents 90% signal and 10% noise, or a signal-to-noise ratio of 9:1. Because the true score is unknown, reliability must be estimated from the sample. Cronbach's alpha[3] is a widely used estimate of reliability [22, p. 299]. The main weakness of this measure is that it assumes that scores are linear and distributed normally. However, scores are not linear [10] and marks are rarely distributed normally in practice. Broadly speaking, the effect of this is to understate the strength of a relationship because the maximum value of the correlation coefficient cannot be reached if these assumptions are not met. The Rasch person separation reliability [23, p. 153] provides a better estimate because it is calculated in a linear metric and is free from distributional assumptions.

It can be noted that person separation reliability was higher than Cronbach's alpha in both scales. This suggests that it would be preferable to use a formal measurement model for the analysis. Nevertheless, since this case study is intended to illustrate the broad method, we proceeded by using the raw scores and treating these as linear measurements. We did this because we believe this makes the method accessible to a wider range of researchers. We give a more formal justification of this decision as follows. The correlation between the linear and raw CICT scales was .935; between the linear and raw further study scales was .964. These are both well above 0.9. Thus, the raw scales clearly capture most

[3] Although widely termed Cronbach's alpha, this was originally derived by Louis Guttman [25] who termed it λ_3.

of the available information in the datasets, with less than a 10% loss of information. Following Agresti [14], it is reasonable to expect that a linear model applied to scores would lead to similar conclusions as a full model. The logic in this study does not require an accurate estimate of absolute effect size, but rather relative estimates. We conclude that the simpler approach is not only more accessible, but would lead to similar conclusions.

5.4 RESULTS

We begin by presenting the mean raw scores of each course across the cohorts in Table 4.

Table 4: Mean scores across all cohorts

Course	Mean	Std. error	N
CICT400	84%	1.3%	112
CICT410	71%	1.5%	98
CICT420	76%	1.9%	114
CICT430	69%	1.2%	121
CS153	62%	2.5%	115
IT181	64%	3.6%	48
PR109	49%	4.2%	37
SE101	69%	4.6%	48

As can be seen in this table, all scores are in the range 60% to 85% with the exception of PR109 which historically has had a low mean score and pass rate. It can also be noted that although the number of students in CS153 is comparable to the CICT numbers, the students in the remaining further study courses are considerably lower. This is because only CS153 is compulsory for the DICT courses, whereas all four are compulsory for the BICT.

We conducted a standard multiple-regression analysis of the contribution of each of the CICT courses to performance as measured by the further study scale. A path diagram of the results is presented in Figure 4.

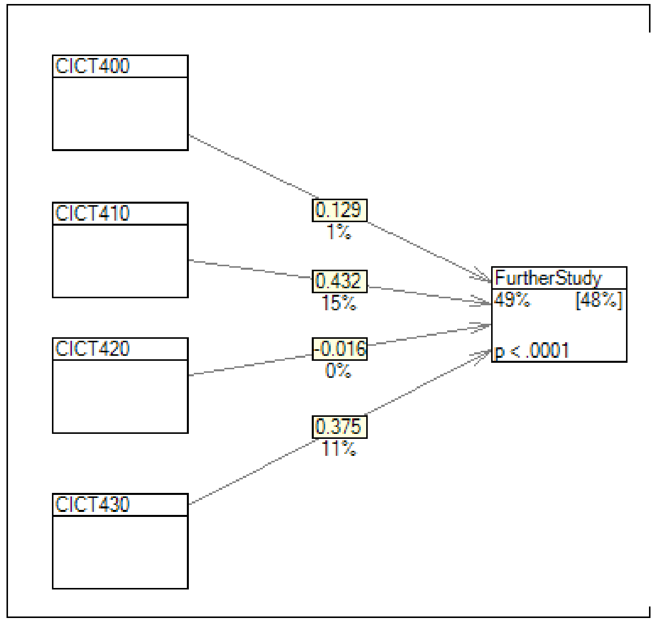

Figure 4: Contribution of CICT courses to subsequent study

In this diagram, titled boxes represent the variables and the arrows represent paths. Each criterion box contains the percentage variation explained (R^2), adjusted R^2 in square brackets, and the probability of a relationship of this magnitude occurring by chance (p value). The beta weight is shown in the boxes on each path and the squared semi-partial correlation is shown underneath. As can be seen, overall, 49% of the variability of marks in further study courses can be explained by overall success in the CICT. Moreover, CICT410 and CICT430 make a unique contribution over and above the commonality of 15% and 11% respectively. We then carried out a multiple regression of the CICT courses as predictors of each of the further study courses. Space does not allow presenting the findings of each course and, thus, we present a composite summary of the results in Figure 5.

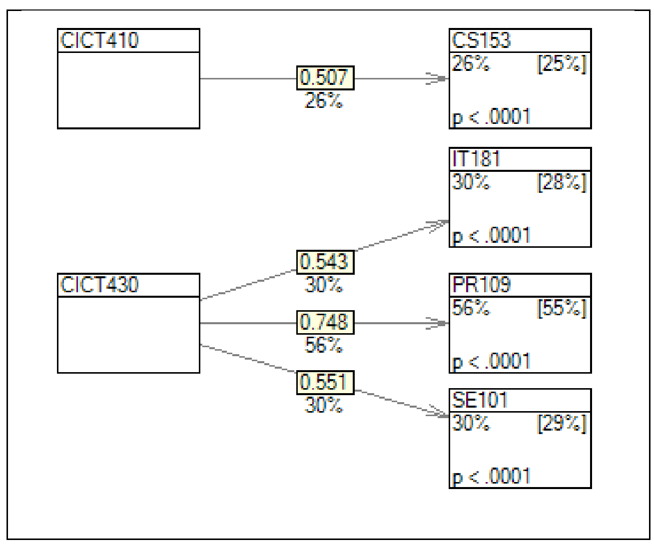

Figure 5: Significant course to course relationships

From this diagram, it can be seen that CICT410 is a strong predictor of success in CS153. Thus CICT410 is, in a sense, the most important course for students who progress to the DICT since this is the only course among those we have researched that is compulsory for the DICT. For the remaining courses, IT181, PR109 and SE101, CICT430 is the only significant predictor among the CICT courses. This suggests that, for students who progress to the BICT, CICT410 and CICT430 are the most important courses from the perspective of preparation for further study.

However, it would be wrong to conclude that there is little value in the remaining courses and we would like to stress the need for care in interpreting results as we mentioned earlier. First, it is important to remember that a programme such as the CICT has purposes beyond the preparation of students and is an exit qualification in its own right. It is reasonable to expect that only some of the skills and knowledge developed will be for preparation for further study and that others will be targeted towards developing the graduate profile for a meaningful exit.

Moreover, we can look at such relationships in two ways. Taking, for example, CICT420 (Information and Communication Skills), we note that this is a professional practice paper which encompasses skills that are essential for everyone in the ICT profession. Thus we should ask, not just how well this prepares students for their next courses in further study, but to what extent these subsequent courses build on the skills and knowledge developed in CICT420.

5.5 CASE STUDY DISCUSSION

This case study investigated and quantified relationships between performance in courses in a level four certificate and performance in courses in two higher level programmes taken subsequently by those completing the earlier certificate. We found that two of the level four courses were strong predictors of subsequent performance, but the other two were not. These findings could be used as an input to any review of the level four CICT to evaluate if they are consistent with the aims of the programme or whether any changes to courses should be considered. Similarly, these findings should be considered by those teaching the later courses from the perspective of asking whether the courses build appropriately on the skills and knowledge established in the earlier programme.

At this point, it is worth taking stock of what has been achieved so far. After controlling statistically for student attributes, marking leniency or severity, commonalities among programs, and elapsed time effects, we have partitioned observed variance into commonalities and unique predictors. This partition has established that two CICT courses are important unique predictors of subsequent performance, whereas the remaining two are not. Any plausible theory must account for why this is so. The review discussed above will produce evidence that either supports the theory of progressive knowledge building or disconfirms it. Either outcome will contribute further to our understanding.

In the introduction, we promised to give a framework for disentangling causal relationships. A cursory review would suggest that there might be other explanations for the pattern observed here. For example, perhaps there is common content between the predictor courses and the criterion. Perhaps students have an aptitude for some topics and it is this that explains the observed relationships. It is at this point that most correlational studies conclude, suggesting that further investigation is needed. However, it is also at this point that the power of the theory-driven approach we are advocating can be seen. Such alternative explanations should be expressed as theories and put to the test. For the sake of discussion, we'll refer to these theories as the *theory of common content* and the *theory of student aptitude*, respectively.

We start with the theory of common content. This can be put to the test in the review we have recommended above. Such a review will either confirm the plausibility of the theory or reject it. In this case, we would expect this theory to be rejected, but the outcome does not matter for the argument we are making. Either outcome will advance our knowledge and is likely to lead to constructive change in how we organise our courses. Testing the theory of student aptitude is more subtle and will be discussed in the next section.

Broadly speaking, we would expect that, from the findings of any individual study, some theories will be rejected and others will remain plausible. This takes our knowledge forward and we have begun to disentangle the possible chains of causation.

6. DISCUSSION AND CONCLUSION

We now address the theory of student aptitude mentioned in the last section. Disentangling causation in this case requires knowledge of the specific courses concerned. As a concrete example, we will suppose that a programme has several programming courses (programming 1, programming 2, etc.) and several networking courses (networking 1, networking 2, etc.). Further, we will assume that programming 1 is a pre-requisite for networking courses. If the theory of student aptitude is correct, we would expect to see strong predictor relationships among the programming courses and among the networking courses, but a weaker relationship between programming 1 and the networking courses. If the theory of progressive knowledge building is correct, we would expect to see a stronger relationship between programming 1 and networking. This difference allows us to put the theories to the test and further disentangle the relationships.

In this paper, we have set out the argument that inference from a correlational study requires a theory-driven approach. However such theory need not be a grand theory of everything. Indeed, the theory can simply be an articulation of experience and tacit knowledge. We have set out here a theory of learning which states that learning happens by building on one's prior experience and, in addition, that successful learning involves three major sets of factors comprising the personal attributes of the learner, the content of a course, and the context set by a teacher. Such a theory seems plausible to us and can be used as the basis for inference. Given such a theory, we can put predicted consequences to the test in a correlational study, thus either disconfirming the theory, or building confidence in its generality. The approach set out in this paper provides one way of putting such a theory to the test.

We have illustrated the approach with a case study. We believe that we have demonstrated in this case study that formal inference from a correlational study is not only possible, but may be useful. However, there are many potential threats to validity in a correlational study such as this. For instance, we make the assumption that performance is accurately captured by the assessment regime. In particular, we note that assessment is likely to be tied to specified learning outcomes and these may not capture the full range of capabilities developed in a course. Such unspecified generic capabilities are likely to relate, in the main, to commonalities between topics (or courses) and, as a consequence, commonality between topics is likely to be underestimated. Nevertheless, because of extensive moderation procedures, we believe that assessment data represent some of the highest quality data available to an educational researcher.

Timperley and Parr [7], challenged us to find ways of making use of available data to inform our teaching decisions and improve our educational performance. Most institutions have a rich source of data available in their assessment regimes but make little use of that data beyond assessment purposes. We have set out a framework in which those data can be used to analyse prerequisite relationships and thus lead to changes which will enhance the student learning experience. Using such naturally occurring data inevitably leads to a correlational approach. Making formal inferences from correlational studies poses significant challenges and requires careful interpretation. However, we believe that the challenges posed by inference from correlational studies should not be used as an excuse for inaction. Accordingly, we advocate wider use of such assessment data in educational research.

7. REFERENCES

[1] K. Popper, The logic of scientific discovery, London: Routledge, 1935/2005.

[2] S. W. Hawking, "Black hole explosions?," *Nature* , vol. 248, no. 5443, p. 30, 1974.

[3] U. Leonhardt, C. Maia and R. Schuetzhold, "Focus on Classical and Quantum Analogs for Gravitational Phenomena and Related Effects," *New Journal of Physics,* vol. 14, no. October 2102, 2012.

[4] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Physics Letters,* vol. 12, no. 2, pp. 132-201, 1964.

[5] A. Del Rosso, "Higgs: the beginning of the exploration," *CERN Bulletin,* vol. 357, no. 47, 2012.

[6] J. A. Sonnenfeld, "Shedding Light on the Hawthorne Studies," *Journal of Occupational Behavior,* vol. 6, p. 125, 1985.

[7] H. Timperley and J. Parr, Using evidence in teaching practice: Implications for professional learning, Auckland, New Zealand: Hodder Moa Beckett Publishers Ltd., 2004.

[8] B. MacMahon, T. Pugh and J. Ipsen, Epidemiologic methods, Boston: Little Brown, 1960.

[9] S. Stevens, "On the theory of scales of measurement," *Science,* vol. 103, no. 2684, pp. 677-680, 1946.

[10] B. Wright and J. Linacre, "Observations are always ordinal; Measurements, however, must be interval," *Archives of physical medicine and rehabilitation,* vol. 70, no. 12, pp. 857-60, 1989.

[11] D. Andrich, "A rating formulation for ordered response categories," *Psychometrika,* vol. 43, pp. 561-73, 1978.

[12] M. Lopez, Measurement of Challenge and Self-Efficacy in Learning [Doctoral Thesis], Perth: Curtin University, 2012.

[13] G. Rasch, Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright., Chicago: The University of Chicago Press., 1960/1980.

[14] A. Agresti, Categorical data analysis, 2nd. ed., Hoboken, NJ: John Wiley and Sons, 2002.

[15] J. Michell, "Quantitative science and the definition of measurement in psychology," *British Journal of Psychology,* vol. 88, pp. 355-383, 1997.

[16] R. Luce and J. Tukey, "Simultaneous conjoint measurement: a new scale type of fundamental measurement," *Journal of Mathematical Psychology,* vol. 1, pp. 1-27, 1964.

[17] G. Karabatsos, "The Rasch model, additive conjoint measurement and new models of probabilistic measurement theory," *Journal of Applied Measurement,* vol. 2, no. 4, pp. 389-423, 2001.

[18] R. Muirhead, Aspects of multivariate statistical theory, Hoboken, NJ, USA.: John Wiley & Sons Inc., 2008.

[19] J. Cohen and P. Cohen, Applied multiple regression/correlation analysis for the behavioral sciences, Oxford, England: Lawrence Erlbaum, 1975, p. 490.

[20] New Zealand Qualifications Authority, *The New Zealand Qualifications Framework [Version 2],* Wellington, NZ: NZQA, 2011.

[21] B. D. Zumbo, A Handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (Ordinal) item scores, Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense, 1999.

[22] L. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika,* vol. 16, no. 3, pp. 297-334, 1951.

[23] B. Wright and M. Stone, Measurement essentials, 2nd. ed., Wilmington, Delaware: Wide Range inc., 1999.

[24] M. Rutter, "Proceeding from observed correlation to causal inference," *Perspectives on Psychological Science,* vol. 2, no. 4, pp. 377-395, 2007.

[25] L. Guttman, "A basis for analyzing test-retest reliability," *Psychometrika,* vol. 10, pp. 255-282, 1945.

,