A Systematic Literature Review of Cognitive Dimensions

Mohammad Hadhrawi
Computer Laboratory
University of Cambridge
mkah2@cam.ac.uk

Alan F. Blackwell
Computer Laboratory
University of Cambridge
alan.blackwell@cl.cam.ac.uk

Luke Church
Computer Laboratory
University of Cambridge
luke@church.name

Abstract

We report a Systematic Literature Review (SLR), exploring the ways that the Cognitive Dimensions of Notations (CDs) framework has been applied since being proposed in 1989. We analyse over 1,600 publications that have cited key references in the CDs literature. Our research questions include 1) whether CDs are used as formative discussion tools or for summative evaluation; 2) which elements of the framework are most widely applied; and 3) the balance between applications and theoretical research contributions.

1. Introduction

Systematic Literature Review (SLR) is a research method introduced in the Software Engineering community, advocated as a rigorous process for meta-analysis of research publications in this field and others (Kitchenham, 2004; Kitchenham et al., 2009, 2010). SLR emphasises the application of quantitative hypothesis-testing procedures to research literature, including: formulation of prior hypotheses; use of replicable search procedures for identifying relevant publications; objective criteria for inclusion (or exclusion) of items in a study corpus; and assessment of hypotheses in a numerical manner (e.g. reporting frequency with which a particular kind of statement is found in the corpus).

In this paper we use SLR to investigate ways in which the Cognitive Dimensions of Notations (CDs) framework has been applied. CDs was first introduced by Green (Green, 1989), then rigorously defined and greatly popularised by Green and Petre (Green & Petre, 1996). Their intention was to create an analytical framework that would allow the usability of notational systems to be compared in a 'broad brush' manner, informed by the body of evidence from earlier empirical studies of programmers, and avoiding the 'death by detail' in which relatively trivial individual programming language features could only be compared by repeating controlled experiments, rather than generalisation from prior understanding of relevant cognitive principles.

The CDs framework was welcomed, both as one of the first comprehensive frameworks synthesising the psychology of programming literature, and also as a source of evidence-based guidance that was presented in a manner accessible to programming language designers. As a result, it has been adopted by programming language design groups in major corporations, and has also been widely influential in the academic community - for example, as the most cited original publication in the Journal of Visual Languages and Computing (Blackwell, 2006). Further overview of CDNs can be obtained from these publications, and from textbook coverage in (Carroll, 2003).

Although apparently widely used in industry, and extended in many ways, our own understanding of how CDs is used has been based largely on anecdotes, and discussions with collaborators. The result of this is that, although there is considerable consensus within the PPIG community regarding the value of CDs, adoption by other communities is not well understood. To the best of our knowledge there has not yet been a systematic review of the use of CDs. In this paper, we address this deficiency by conducting a systematic literature review of the papers that cite CDs.

2. Methodology

Following the guidelines proposed by Kitchenham (Kitchenham, 2004; Kitchenham et al., 2009, 2010), we conducted a systematic literature review to explore the ways that CDs framework are applied. We are

interested in analyzing 1) if CDs framework is used for evaluation or discussion, 2) which elements are widely used, and 3) How the cognitive dimensions of notation framework shaped the current literature.

2.1. Research Questions

During this research, we investigated the following research questions:

RQ1. Are CDs used as formative discussion tools or for summative evaluation?

RQ2. Which elements of the framework are most widely applied?

RQ3. How have CDs publications shaped the research literature?

We used quantitative and qualitative methods in our analysis to investigate these questions. To answer **RQ1**, we introduced the following subquestions:

RQ1.1 Did the citation occur only in a context that discussed evaluation of a completed design?

RQ1.2 Did the citation occur in a context where design properties were being discussed in the light of cognitive dimensions?

Similarly, we wanted to investigate the following subquestions to answer **RQ2**:

RQ2.1 Which of the dimensions proposed in (Green & Petre, 1996) are most widely applied?

RQ2.2 Have the newer dimensions proposed in (Blackwell et al., 2001) been applied?

RQ2.3 Which of the distinctive activities formulated in (Green & Petre, 1996) are applied?

RQ2.4 Which of the other CDs framework elements outlined in (Carroll, 2003) are most widely applied?

For **RQ3** we wanted to investigate the following subquestions:

RQ3.1 Are CDs used in other ways, beyond discussion and evaluation of notational systems?

RQ3.2 What general scientific principles have been adopted from the CDs literature?

2.2. Search Process

We identified nine articles (listed in Table 1) to be the root publications for the cognitive dimension domain. Then, we collected all papers that cite any of these publications.

To create our corpus, we used Google Scholar to find metadata for papers that referenced a paper in the root collection. We generated a list that includes publication titles, number of citations, authors, date of publication, and whether the result has a direct PDF file link, or not. Then, we removed all non-English titles and duplicates from this list. Next, we downloaded all the PDF files from the filtered list. For publications that did not have a direct PDF link to download, we manually entered the full publication title as a Google search, reviewing the first two pages of search results for any site that would provide a downloadable PDF file of that publication. We used University of Cambridge library credentials to obtain a PDF from any publisher providing a valid institutional subscription. Where PDFs were paywalled (for example, some Springer conference proceedings), we did not proceed further.

For each downloaded PDF, we searched for specific strings to assess whether to include or exclude that publication. The remaining PDF files were searched for strings “green” or “cognitive", and if the matched text is not a reference to cognitive dimensions (e.g., the colour green, or reference to cognitive psychology), we continued searching for repeat alternatives until the end of the file. If the first occurrence of “cognitive dimensions”, “green” or “cognitive” occurred in the bibliography, we noted the citation reference, usually an index number or BibTeX code, and searched the paper for this citation.

ID	Name
root01	Usability analysis of visual programming environments: a 'cognitive dimensions' framework (Green & Petre, 1996)
root02	Cognitive dimensions of notations (Green, 1989)
root03	Notational systems - the cognitive dimensions of notations framework (Blackwell & Green, 2003)
root04	Cognitive dimensions of information artefacts: a tutorial (Green & Blackwell, 1998a)
root05	Design for usability using Cognitive Dimensions (Green & Blackwell, 1998b)
root06	A Cognitive Dimensions questionnaire optimised for users (Blackwell et al., 2001)
root07	Cognitive dimensions of notations: Design tools for cognitive technology (Blackwell et al., 2001)
root08	The cognitive dimension of viscosity: a sticky problem for HCI (Green, 1990)
root09	Instructions and descriptions: some cognitive aspects of programming and similar activities (Green, 2000)
root10	Ten years of Cognitive Dimensions in visual languages and computing: Guest Editor's introduction to special issue (Blackwell, 2006)

Table 1 – Source/Root Publications. Note that root04 and root05 are the same publication but cited differently.

Search for index number 99 was speeded up by initially searching for "[99]", "[99", "99]", or "99,", in order to avoid occurrences of the same digits in a year, page number, data values etc. If one of the root papers had appeared in the bibliography, but no citation was found by the above search procedure, this was coded as "no citation". Note that this does not guarantee that there was no citation, but simply that, if a citation did occur, it was phrased in a manner that did not match any of the above search strings. It is possible that for some PDF files, this resulted from OCR processing being more accurate for the bibliography than for the body text (for example, body text was sometimes incorrectly encoded as image regions).

2.3. Inclusion and Exclusion Criteria

Initially, we included all publications which cited any of the nine root articles. However, we had to exclude a subset of this collection due to: a) *publications not written in English*, b) *duplicate publications*. To remove non-English publications, we reviewed all titles manually, excluding any article with a non-English title. Also, we did a sanity check of titles by searching for Unicode characters outside the standard English character set. Using this new filtered collection, we removed remaining duplicates manually by sorting titles alphabetically, and by running a string comparison script.

In addition to duplicates and non-English publications, we excluded all c) *publications without a downloadable PDF file*. From our manual search, we either could not find a downloadable PDF file in the first two pages of the Google search results, or the publication was not accessible due to licensing and purchase requirements.

Furthermore, we excluded all d) *PDF files that do not support text search*. Usually this is because PDF files are image scans of old papers, or occasionally because OCR was incomplete or faulty.

Finally, e) *publications without specific string occurrences* are excluded. These strings are: "cognitive dimensions", "green" (since Thomas Green was a co-author on all root papers) or "cognitive" (in this order).

2.4. Data Analysis

To investigate and answer our research questions, we conducted a quantitative, contrast and qualitative analysis.

2.4.1. Quantitative Analysis

To conduct a quantitative analysis on our corpus, we converted all PDF files to text. Then, we searched and counted the frequency of occurrence of certain keywords in each PDF. We separately considered four different sets of keywords: first, the original 'core' set of cognitive dimensions (RQ2.1); second, the set of 'new' dimensions (RQ2.2); third, notational activities (RQ2.3); and fourth, other CDs framework components (RQ2.4). These keywords along with their regex search terms are listed in Table 2.

2.4.2. Contrast Analysis

To support our quantitative analysis, we repeated the quantitative analysis process from the previous section on the proceedings of CHI 2016 (Kaye & Druin, 2016) and CHI 2017 (Mark & Fussell, 2017). The CHI dataset can be expected to slightly over-estimate the usage of cognitive dimensions terminology, when compared to other writing in English.

The occurance frequencies in the CHI dataset are used as contrast material for the occurence frequencies in the papers that cite cognitive dimensions. If a word occurs frequently in the CHI dataset it is excluded from further analysis or reporting in the quantitative sections of this paper as it would not be possible to determine whether it was used frequently because of its relevance to cognitive dimenions or a more general term of art.

2.4.3. Qualitative Analysis

In this section we describe the coding scheme for the qualitative analysis. The primary qualitative coding procedure was based on the text context in which the citation appeared. In some papers, there were multiple citations, in which case each occurrence was inspected before making the coding decision. The coding decision procedure meant that there were no cases in which multiple citations resulted in ambiguity of the coding decision.

- If citations occurred only in a context that discussed evaluation of a completed design (including the design of a notation such as UML, or the design of an interactive system), then this paper was coded as "evaluation".
- If citations occurred in a context where design properties were discussed in the light of cognitive dimensions, this was coded as "discussion". Note that, in some cases, discussion occurred within a context that might be described as formative evaluation of alternative design options. These cases were coded as discussion, if CDs were being used as a discussion vocabulary within a design process rather than for assessment of a completed design. In many cases, only a small number of dimensions were mentioned - we discuss this further in the results section.
- If the context of the citation did not mention any cognitive dimensions, did not discuss a design in terms of CDs, and did not mention evaluation, it was coded as "only cite". in these cases, the authors were citing one of the root papers for a purpose unrelated to the main objective of CDs, for example as a reference to support mention of visual languages.

For those cases where a root paper was cited, but with no discussion of CDs, the apparent topic was recorded. An open coding approach was taken, in order to allow summary frequencies of the purposes for which citations were made. In cases where the authors used the citation to support a statement that directly contradicted the core contributions of CDs (for example, a claim that visual languages are always superior to text, or that text is always superior to visual languages), the complete text of the statement was copied for later analysis.

3. Results

Our search process resulted in an initial count of 2,555 articles. After cleaning the data, removing duplicates, non-English titles and non-downloadable PDFs, we ended up with 1,638 articles that we run our analysis on.

3.1. Quantitative Results

We address the subquestions of RQ2 through analysis of keyword frequencies. Each plot shows a histogram of the number of publications that mention each keyword a particular number of times. For each keyword, the mode is 1 (i.e. in most cases, the keyword is only mentioned once in the publication). However, there are a few publications that repeat a keyword many times - sometimes more than 30 times.

3.1.1. Core Dimensions

*Figure 1 – CDs Framework Core Dimensions Histogram Comparison. Each histogram represents a dimension with the dimension name as the title, and the regex search term as the subtitle. In the parentheses (**a** in **b**), **a** is the total number of occurrences , and **b** is the total number of publications where the words occurred. The x-axis is the word occurrences, and the y-axis is the number of publications.*

Figure 1 shows the frequency distribution charts of the core dimensions. Words that had high frequency in the contrast materials as descrived in section 2.4.2 where excluded. These were "abstraction", "consistency" and "visibility". "viscosity" was marginal, having slightly higher frequency in the contrast material than the other terms here, so it should its high occurance here may be partially accounted for it being a semi-common word in the broader field beyond CDs.

From these charts it appears that "Viscosity" dimension has the highest number of counts and appeared in more articles than other core dimensions (3,116 counts in 558 articles). Whereas "Provisionality" has the lowest number of counts and appeared less compared to other core dimensions (416 counts in 140 articles).

From these results, we found that some articles have a specific dimension keyword that occurred over 40 or 50 times. These instances were used to introduce the dimensions keyword as part of new dimensions/framework, or the article is primarily focused on that specific dimension. For example, one article used the word "provisionality" 45 times. This article used this dimension and introduced it as part of communicative dimensions. Another article used "premature commitment" 50 times, which focused

primarily on Premature Commitment dimension.

When dimensions keywords occurred more than 20 times, and in some cases over 10 times, in a single article, the article would use the dimension for evaluation. For example, when the dimension "closeness of mapping" occurred over 10 times in a single article, the article describes this dimension and use it for evaluation or comparison.

3.1.2. New Dimensions

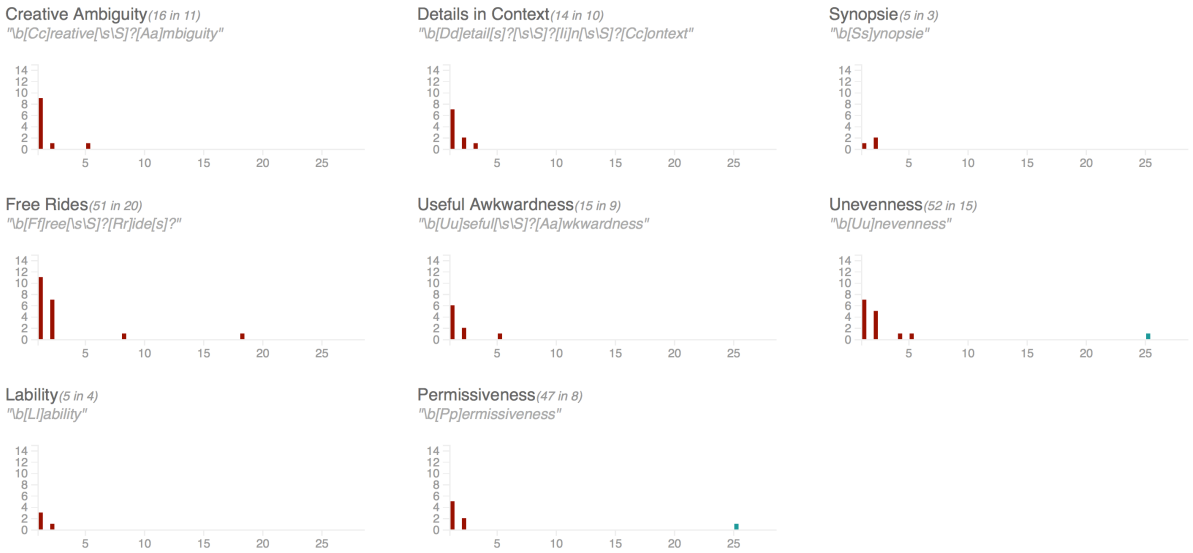

Figure 2 – CDs Framework New Dimensions Histogram Comparison

The new dimensions (RQ2.2) appeared far less frequently than the core dimensions, as seen in figure 2. As with the core dimensions results, we removed two keywords that have high frequency in the contrast material:"indexing" and "specificity". However, an interesting finding from our analysis appears in the keyword "permissiveness". The results shows an occurrence of 47 times over eight publications of which one of these publications used the word 42 times. When we investigated that particular publication, we found that the publication is introducing that new dimension to the framework.

3.1.3. Activities

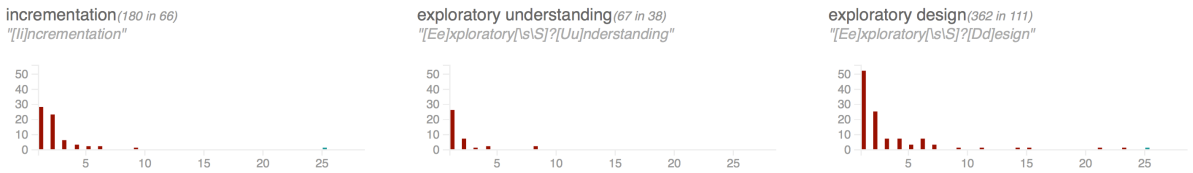

Figure 3 – CDs Framework Activities Histogram Comparison

As in previous sections, we excluded the two CDs activities because of their frequency in English: "modification" (2,561 counts in 420) and "transcription" (478 counts in 138 articles).

From figure 3, the CDs framework "incrementation" activity occurred 180 times in 66 articles. One of these articles used the "incrementation" term 45 times. From our investigation, it appeared that the author is describing the different CDs framework activities in every dimension.

An interesting finding is the word count of the CDs framework activity "exploratory design", which occurred 362 times in 111 articles.

3.1.4. Other Framework Components

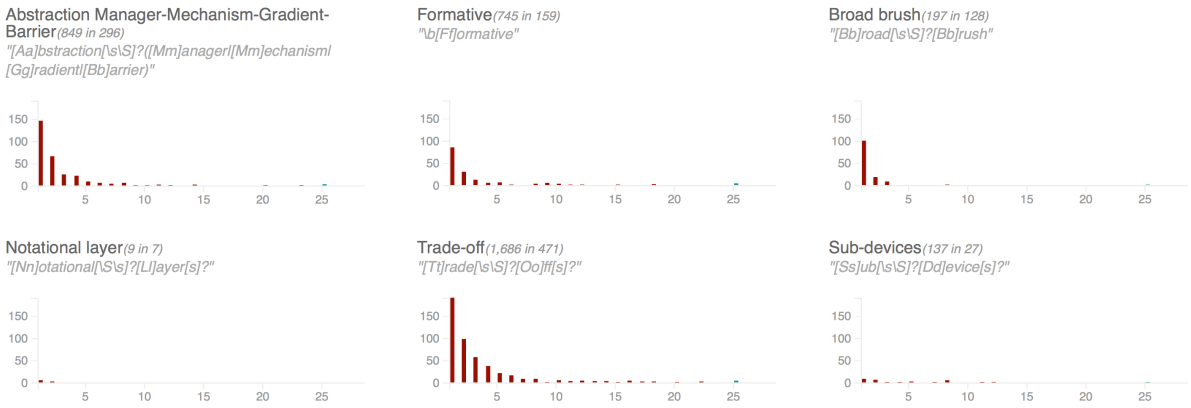

Figure 4 – Other CDs Framework Components Histogram Comparison

Figure 4 compares keywords related to other aspects of the CDs framework. Once again, we exclude terms that frequent in the contrast material: "evaluation" and "notation". The component "Trade-off" has the highest number of occurrences, which occurred 1,686 times in 471 articles. However, based on random sampling of these articles, it seems that the word is not being used primarily in association with the CDs framework.

"Notational layer", which occurred nine times across seven articles is the least used component. "Abstraction Manager-Mechanism-Gradient-Barrier" component has a relatively high occurrences (849 times in 296 articles). This could be due to our query for that aggregate the results of the four keywords together "Abstraction Manager", "Abstraction Mechanism", "Abstraction Gradient" and "Abstraction Barrier".

Investigating the "sub-devices" component, which appeared 1,237 times in 27 article, we found that when the word occurred over 10 or 20 times within one article, that article used the component in an evaluation context.

3.2. Qualitative Results

3.2.1. Further exclusions

71 papers were removed from the corpus during the qualitative coding process. 13 of these were not in English, but had not been detected by our initial filter. The remainder were documents that are indexed by Google Scholar, but were not academic publications, and made no reference to CDs. 12 papers were image scans with no searchable text. 50 further papers had no identifiable citation to any of the root papers, following the procedure specified above. Furthermore, 56 publications were determined to be duplicates, with titles that were identical or close to identical.

3.2.2. Publications *about* CDs

90 publications in our dataset were classified as contributions to, or extensions of, the CDs framework itself. This count includes the original root papers (for completeness of the dataset in relation to citation metrics), presentations of the framework to other disciplinary audiences, methodological contributions proposing ways of using cognitive dimensions in applications or research, theoretical extensions and alternatives to the CDs approach such as the Tangible Correlates (Edge & Blackwell, 2006) or Ontological Sketch Modelling (Blandford & Green, 1997), and other frameworks inspired by the CDs approach such as Communicative Dimensions (Hundhausen, 2005), Collaborative Dimensions (Bresciani, Blackwell, & Eppler, 2008), "Physics of Notations" (Moody, 2009), and the Patterns of User Experience (Blackwell & Fincher, 2010). We do not further analyse these - we presume that much of this work is familiar to the specialist audience, and our main concern here is to analyse how CDs has been applied, rather than the history of how it has been, and continues to be, developed.

3.2.3. Use of CDs for summative evaluation

A total of 211 papers met the criteria defined above for a publication that reports using CDNs to evaluate a design, rather than as a discussion tool informing the research or design process. One of these claimed that the evaluation had been "formative", but was reported as a summative contribution (within a project validation phase). 22 papers expressed an intention to evaluate a system using CDs, but reported this in a "future work" section, meaning that it may or may not have occurred. One reported that the authors had intended to do an evaluation, but did not. 5 papers simply stated that the system had been evaluated using CDNs, but provided no details of what this had involved, or what the outcome had been. 25 of the evaluation papers stated that they had used the CDs questionnaire. These involved a fairly wide variety of questionnaire respondents, ranging from members of the research team to experimental participants. Two papers reported more "objective" evaluation using CDs - in one case as the basis of an experiment design, and in another as a numerical score of the number of dimensions that had been satisfied.

3.2.4. Use of CDs for design discussion

585 publications cited the root papers in order to discuss cognitively-relevant system properties, in the manner that may be considered as closest to the original advocacy in papers by Green, and by Green and Petre. In 16 of these cases, the authors expressed an intention to engage in CDs analysis, but this was stated in a "future work" section, with no further content. One paper simply stated that discussion had taken place, but gave no further information. 11 papers used the CDs questionnaire to structure discussion, among members of a design team or with potential users, while another described CDs as "a discount usability method, [that] can be used without users".

The remaining 555 papers in this category simply go about the business of discussing specific dimensions, considering desirable attributes of notational systems for particular applications, in the way that a PPIG audience might expect. This discussion involved a wide range of different degrees of detail. Some papers were devoted to comprehensive CDs analysis of a particular application area, including consideration of activities, profiles, relevant dimensions and so on. However, this was relatively rare, as indicated by findings in the quantitative analysis above. More often, authors simply named specific dimensions as being relevant to their work, and then proceeded to discussion of design factors that would be influenced by that dimension. The relative frequency of dimensions chosen has been discussed in the quantitative analysis above.

One distinctive characteristic was the large number of these papers, 208, in which a single dimension was the main focus of attention. Different dimensions were relevant on different occasions, as reflected in the numerical distributions reported above.

15 papers use the CDs as a coding frame for qualitative analysis. In two of these cases, the qualitative data was collected using the CDs questionnaire as a prompt to participants, while the remainder collected data in other ways, including interviews, focus groups, Wizard of Oz and think-aloud studies. We consider these cases as a discussion tool for researchers, despite the fact that a specific design may not be under discussion.

3.2.5. Further citations of the CDs framework

Finally, 521 of the publications in our corpus cite one of the root papers, but do not engage either in discussion or evaluation of notational systems. In many cases, these are simply "completeness" citations, in which the authors have presumably found it necessary to report that the CDs framework exists, but see no pressing need to apply it in their own research. This applies to 260 of the cases in this category.

3.2.6. Open coding of other citations

Slightly more interesting are the publications that cite one of the root papers, but for other purposes beyond application or extension of CDs. As described in the methods section, we carried out further thematic open coding of these 261 papers.

We found 46 papers cite the Green and Petre paper (Green & Petre, 1996) as reference for the concept of a visual language. A further 4 papers cite it as a reference to visualisation or representation, 7 as

a reference to domain-specific languages, 4 to dataflow languages, 1 to high-level languages and 4 as a reference to user interfaces in general. The paper is also used as a citation for specific languages that appear as case study examples: 2 citations as a reference to spreadsheets, 4 to LabView, and 1 to Prograph.

27 publications cite CDs papers as general references to cognitive psychology, including the nature of cognition (10), learning (7), cognitive load (2), as well as perception, understanding, working memory, tacit knowledge, implicit information, external representation etc. 18 papers cite the Green and Petre paper as a reference to support the concept of usability in general, and 1 as a reference to the field of HCI. 5 papers cite a CDs paper as a reference for psychology of programming or PPIG.

Specific kinds of programming activity are another focus of citations, with 6 papers referring to CDNs as evidence for exploratory design, 4 for opportunistic planning, 2 for plan composition, 2 for direct manipulation, 1 for information seeking, 1 for searching, 1 for diagram editing, 1 for program comprehension, 1 for modularisation, 2 citing novice/expert differences, and 3 for programming in general.

35 publications cite a CDs paper in support of earlier work by Green and colleagues. 19 of these address the problem of superlativism - the assumption that any one programming language could be best for all purposes - while 16 report the "match/mismatch" finding, that notations are effective in situations where they match the structure of the task.

CDs is reasonably often cited to illustrate a general analytic strategy, including an emphasis on notation (11), on information artefacts and representation (7), or on editing environments (1) as well as the way these concerns are addressed in design processes (6) with the consequent need to acknowledge trade-offs in design (5). Three papers refer to CDs to demonstrate the value of design vocabularies as a research contribution, while another two cite it as an example of a research framework that is open and extensible. Two papers (mistakenly) cite a CDs publication as the source for "Physics of Notations".

4. Discussion and Future Work

From our search process, we found that some articles were indexed by Google Scholar, but were not academic papers or did not make reference to CDs. We attribute these to data quality issues in Google Scholar, or the way in which we were using it. In future work conventional citation services, although having less complete coverage than Google Scholar, could be used to test whether bugs resulted in some publications not being found in our search procedure.

For future filtering, a more sophisticated filter could use an edit-distance metric to check for duplicate titles. However, we note that a number of authors publish very similar papers with slightly different titles. This is, unfortunately, common academic practice. In this analysis we have followed academic convention by counting these as separate publications if they appeared in separate venues. However, it is an issue that should perhaps be addressed more rigorously in future SLR studies.

We mentioned that we excluded publications that are images scans with no searchable text. We would like to go over the image scans articles and code them manually in future work.

From the quantitive results, it appears that the "new" dimensions suggested as extensions beyond the set described in the earliest CDs publications appears far less frequently (RQ2.2). This suggests that many authors regard the original set as a complete basis for evaluation, despite the stated intention that the framework should be considered as extensible. We noted during this analysis that articles refer to exploratory design activity more often than other activities (notwithstanding our exclusion of very common terms). This reflects the fact that exploratory design has not been a common focus of attention in software engineering research, meaning that CDs has drawn attention to a phenomenon that was previously neglected.

From the qualitative results, more often, evaluation using CDs was complementary to a more conventional user study, with the results of the two investigations being aggregated for summative evaluation of a prototype system. This approach was particularly associated with certain research groups.

In our analysis of the use of CDs for design discussion, the use of CDs to draw attention to specific properties of interactive systems is very much consistent with Green's original observations and intent, as in the highlighting of "viscosity" as a specific dimension that could become a focus of greater design attention in future (Green, 1990). As seen in the quantitative analysis, the CDs literature has been influential in promoting a variety of specific design concerns for notational systems, including secondary notation, and in particular closeness of mapping, which has been received by the research community as a property that is especially valuable and significant for the design of domain-specific notations.

As mentioned in the qualitative results section, some publications were coded as using CDs for discussion, despite the fact that no specific designed artefact is discussed. However, the kinds of activity resulting in qualitative data were all broadly situated within a design research context, meaning that the authors' findings might in future be applied in design work. The value of a discussion tool here is in focusing analysis by the researchers toward design concerns that are likely to be relevant in notational systems.

The results indicate a significant number of publications that cite the root papers without engaging in discussion or evaluation of notational systems. This is a typical academic convention, that work may be cited due to its reputation, rather than because of a specific influence on the citing work. This is a cautionary observation, not only for members of the PPIG community who hope to have some influence on future technical practice and might use CDs as a benchmark of their highest aspirations, but also for those who might be tempted to regard citation metrics as strong evidence of scientific value.

As the results suggested in the open coding of other citations, there are a number of papers that cited the Green and Petre paper (Green & Petre, 1996) as a reference for the concept of a visual language. Since it was published in the Journal of Visual Languages and Computing, presumably any paper in this journal would have done - or indeed a citation to the journal itself - but this paper was taken to be in some way representative of the whole field. In addition, the Green and Petre paper offers a rich selection of case studies, and is occasionally cited by authors who have no interest in CDs, but simply need a source for one of these. It is thus cited as reference for the rocket simulator program, for logic gates, or paperless interfaces. Some are more puzzling, for example citing a CDs publication as a reference for XML pairing, use of space, of metaphor, or for 'non-linear art'. Some of these may have been referencing errors, or perhaps more innovative chains of reasoning and connection that, while clear in the author's mind, have not been fully communicated to the reader.

From the 35 publications that cite the CDs in support of earlier work by Green and colleagues, some authors appear to have misinterpreted these central points of the framework, for example "In accordance with Green et al., we agree that graphical representations are, in most cases, inherently superior to textual representations" or, as another author states in direct contradiction, "But Visual Programming Languages did not work well in practice as diagrams cannot be as expressive as text". Although these quotes are relatively extreme, several other papers cite CDNs in passing as a reference to support either the claim that text is inherently superior to graphics, or that graphics are inherently superior to text.

From our qualitative results, there are a small number of papers that reflect on design theory, including the need for frameworks, the role of evaluation etc. Unfortunately, a small number of these also seem to ignore or misunderstand the key concerns of CDs, contradicting the central message, as in "Generally, the Cognitive Dimensions framework cannot be used to directly guide us in creating visually-enhanced notations for code."

5. Conclusions

This paper has presented the results of a systematic literature review, studying applications of the Cognitive Dimensions of Notations framework, as discussed in over 1,600 publications. The CDs framework is very widely cited, and considered influential, yet much of the literature draws on CDs in ways that are different to the original intentions of the framework developers.

We did find a significant number of papers that use CDs as a discussion tool, in order to draw attention

to cognitively-relevant aspects of system design. But in many cases, the discussion focused on a single dimension, meaning that the role of CDs has been to alert language designers to a particular property not previously understood, rather than introducing new sophistication to the design process mnore broadly.

There are many richer discussions of CDs properties, but these are often used in a purely summative way, after the design is completed, rather than as intended to inform trade-offs and design manoeuvres during an ongoing process of design guided by research. In some cases, even evaluation appears to be an afterthought, with authors expressing their intention to evaluate their work in the future, presumably placing higher priority on publication of their design proposals than on evaluation of the usability benefits they hope these proposals might deliver.

As one of the most highly cited references in Psychology of Programming, it is also interesting to note that the most popular CDs publication is taken as a single representative source that is assumed (from the perspective of software engineering researchers) to represent everything that is known in our field, or in the fields of visual languages, usability, HCI, or cognitive science.

When a paper becomes so widely cited, it is inevitable that it should also become misunderstood, and it is not unusual for researchers to cite this paper in support of claims that are opposite to the actual scientific findings. These practices suggest a degree of laziness in contemporary peer review, and perhaps a lack of rigour in fields such as Software Engineering and Information Systems. Although our own field often values innovation and insight beyond the incremental observation of controlled experiments, we should be alert to the potential for misunderstanding, amongst audiences who may be unfamiliar with design-oriented research and broad-brush analysis.

6. References

Blackwell, A. F. (2006). Ten years of cognitive dimensions in visual languages and computing: Guest editor's introduction to special issue. *Journal of Visual Languages & Computing*, *17*(4), 285–287.

Blackwell, A. F., Britton, C., Cox, A., Green, T. R. G., Gurr, C., Kadoda, G., ... Young, R. (2001). Cognitive dimensions of notations: Design tools for cognitive technology. In *Cognitive technology: Instruments of mind* (pp. 325–341). Springer.

Blackwell, A. F., & Fincher, S. (2010). Pux: Patterns of user experience. *interactions*, *17*(2), 27–31.

Blackwell, A. F., & Green, T. R. G. (2003). Notational systems–the cognitive dimensions of notations framework. *HCI Models, Theories, and Frameworks: Toward an Interdisciplinary Science. Morgan Kaufmann*.

Blandford, A., & Green, T. R. G. (1997). Osm: an ontology-based approach to usability evaluation. In *Proceedings of workshop on representations.*

Bresciani, S., Blackwell, A. F., & Eppler, M. (2008). A collaborative dimensions framework: Understanding the mediating role of conceptual visualizations in collaborative knowledge work. In *Proceedings of the 41st annual hawaii international conference on system sciences (HICSS 2008)* (pp. 364–364).

Carroll, J. M. (2003). *HCI models, theories, and frameworks: Toward a multidisciplinary science*. Morgan Kaufmann.

Edge, D., & Blackwell, A. F. (2006). Correlates of the cognitive dimensions for tangible user interface. *Journal of Visual Languages & Computing*, *17*(4), 366–394.

Green, T. R. G. (1989). Cognitive dimensions of notations. *People and computers V*, 443–460.

Green, T. R. G. (1990). The cognitive dimension of viscosity: A sticky problem for HCI. *Human-Computer Interaction*, 79–86.

Green, T. R. G. (2000). Instructions and descriptions: some cognitive aspects of programming and similar activities. In *Proceedings of the working conference on advanced visual interfaces* (pp. 21–28).

Green, T. R. G., & Blackwell, A. F. (1998a). Cognitive dimensions of information artefacts: a tutorial. In *BCS HCI conference* (Vol. 98).

Green, T. R. G., & Blackwell, A. F. (1998b). Design for usability using cognitive dimensions. In *Tutorial*

session at british computer society conference on human computer interaction HCI (Vol. 98).

Green, T. R. G., & Petre, M. (1996). Usability analysis of visual programming environments: a 'cognitive dimensions' framework. *Journal of Visual Languages & Computing*, *7*(2), 131–174.

Hundhausen, C. D. (2005). Using end-user visualization environments to mediate conversations: a 'communicative dimensions' framework. *Journal of Visual Languages & Computing*, *16*(3), 153–185.

Kaye, J., & Druin, A. (2016). *CHI '16: Proceedings of the 2016 CHI conference on human factors in computing systems.* New York, NY, USA: ACM.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, *33*(2004), 1–26.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology*, *51*(1), 7–15.

Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering–a tertiary study. *Information and Software Technology*, *52*(8), 792–805.

Mark, G., & Fussell, S. (2017). *CHI EA '17: Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems.* New York, NY, USA: ACM.

Moody, D. (2009). The "physics" of notations: toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering*, *35*(6), 756–779.

	Keyword	Regex Search Term
Core Dimensions	Viscosity	"[Vv]iscosity"
	Visibility	"[Vv]isibility"
	premature commitment	"[Pp]remature[\s\S]?[Cc]ommitment"
	hidden dependencies	"[Hh]idden[\s\S]?[Dd]ependencies"
	role-expressiveness	"[Rr]ole[\s\S]?[Ee]xpressiveness"
	error-proneness	"[Ee]rror[\s\S]?[Pp]roneness"
	abstraction	"[Aa]bstraction"
	secondary notation	"[Ss]econdary[\s\S]?[Nn]otation"
	closeness of mapping	"[Cc]loseness[\s\S]?[Oo]f[\s\S]?[Mm]apping"
	consistency	"[Cc]onsistency"
	diffuseness	"[Dd]iffuseness"
	hard mental operations	"[Hh]ard[\s\S]?[Mm]ental[\s\S]?[Oo]peration[s]?"
	provisionality	"[Pp]rovisionality"
	progressive evaluation	"[Pp]rogressive[\s\S]?[Ee]valuation"
New Dimensions	Creative	"\b[Cc]reative[\s\S]?[Aa]mbiguity"
	Specificity	"\b[Ss]pecificity"
	Detail in context	"\b[Dd]etail[s]?[\s\S]?[Ii]n[\s\S]?[Cc]ontext"
	Indexing	"\b[Ii]ndexing"
	Synopsie	"\b[Ss]ynopsie"
	Free rides	"\b[Ff]ree[\s\S]?[Rr]ide[s]?"
	Useful awkwardness	"\b[Uu]seful[\s\S]?[Aa]wkwardness"
	Unevenness	"\b[Uu]nevenness"
	Lability	"\b[Ll]ability"
	Permissiveness	"\b[Pp]ermissiveness"
Activities	transcription	"[Tt]ranscription"
	incrementation	"[Ii]ncrementation"
	modification	"[Mm]odification"
	exploratory design	"[Ee]xploratory[\s\S]?[Dd]esign"
Others	Evaluate, evaluation...etc.	"[Ee]valuat[\w]+"
	abstraction manager	"[Aa]bstraction[\s\S]?([Mm]anager|[Mm]echanism|[Gg]radient|[Bb]arrier)"
	abstraction mechanism	"[Aa]bstraction[\s\S]?([Mm]anager|[Mm]echanism|[Gg]radient|[Bb]arrier)"
	abstraction gradient	"[Aa]bstraction[\s\S]?([Mm]anager|[Mm]echanism|[Gg]radient|[Bb]arrier)"
	abstraction barrier	"[Aa]bstraction[\s\S]?([Mm]anager|[Mm]echanism|[Gg]radient|[Bb]arrier)"
	notation	"[Nn]otation"
	formative	"[Ff]ormative"
	broad brush	"\b[Bb]road[\s\S]?[Bb]rush"
	notational layer	"[Nn]otational[\s\S]?[Ll]ayer[s]?"
	trade-off	"[Tt]rade[\s\S]?[Oo]ff[s]?"
	sub-devices	"[Ss]ub[\s\S]?[Dd]evice[s]?"

Table 2 – Quantitative analysis keywords and their regex search terms