B.Sc. PROJECT REPORT

2017

Development of the graphical representation of hierarchical task analysis

ANDREW REECE

B416391

Submitted in part-fulfilment of the requirements of the examination

for the B.Sc. (Honours) in Ergonomics

Abstract

Aims

Development of the graphical notation of hierarchical task analysis (HTA) has not kept pace with the tabular notation, despite advances in graphic communication theories since its inception. The project aimed to rectify this with a mixed-methods approach: contextualise practitioner use, revise the notation, evaluate revisions empirically and based on stakeholder feedback.

Methods

Study 1: 22 HTA practitioners were asked 12 open questions on their use of HTA. The responses were inductively analysed for common themes.

'Physics of notations' principles were applied to the traditional notation for initial improvements. Two new notations were created.

Study 2: 6 participants unfamiliar with HTA notation were shown both new notations and the traditional in a random order, which they interpreted. A 4th revised notation was created based on their feedback.

Study 3: 29 students (recent HTA trainees) were presented with questionnaires and traditional or revised notation task diagrams. They were timed responding to surface-level-comprehension questions, and they provided multiple subjective ratings.

Study 4: 23 HTA practitioners provided feedback on the notational revisions.

Results

Graphical HTAs are used for communication with ergonomists and non-ergonomists.

Complete novices prefer a graphical plan representation and the revised hierarchy notation.

Students answered more questions correctly ($p=.016$) and took less time per correct answer ($p=.031$) with the revised notation. Improvements for individual questions were only significant for task occurrence ($p=.003$) and - for males - hierarchy ($p=.033$). The revised notation was rated more complex ($p<.001$). Gender differences were noted for accuracy and preference.

Practitioners considered many revisions to be improvements. Suggestions, including for reducing the visual complexity, were made.

Conclusions

Graphical HTAs should indicate non-definite tasks; dashed lines represent this well. The revised notations for hierarchy and plan sequencing appear preferable to the traditional, however additional investigation is recommended. Further improvements and research directions are discussed.

Abstract Word Count: 300/300

Report Word Count: *7,497 / 7,500*

Acknowledgements

I would like to thank my supervisor Simon Hodder for regularly giving up his time to talk through half-thought-out ideas, propose suggestions, and provide course corrections when I looked in danger of getting lost in tangentially relevant processes.

My thanks also go to the many ergonomics practitioners who took time out of their busy schedules to share their expertise with me, and the other participants, without whom the project would not have been feasible at all.

The friends and family members who proofread the report - possibly not realising quite how long and esoteric it would be when they offered to do so - are given my sincere thanks.

Table of Contents

List of Figures

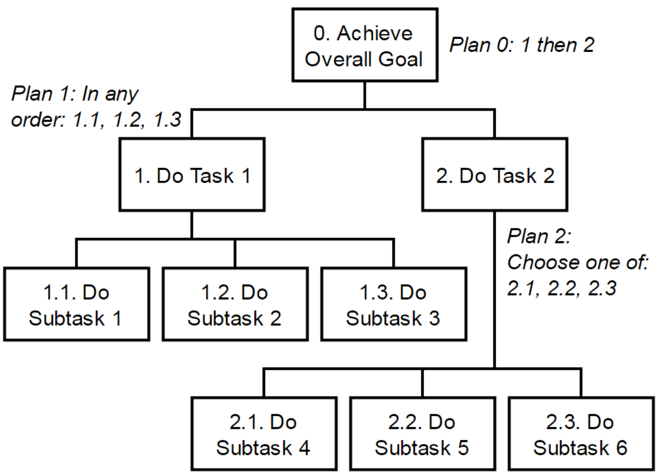
List of Tables

1. Introduction

Hierarchical Task Analysis

Overview

Hierarchical task analysis (HTA) provides a method for logically decomposing an overall task recursively into sub-tasks until there is sufficient detail for the user's particular purpose. This goes together with a 'plan' that describes the context and sequencing of sub-tasks. HTA is represented in two ways:a tabular format that considers each task individually, and a graphical format that represents tasks as boxes in a tree structure (Figure 1). Both are accompanied by a textual plan. HTA is "the best known task analysis technique" (Kirwan and Ainsworth, 1992:396) and a "central approach" (Stanton, 2006:55) in ergonomics research.



Super-ordinate number	**Goal** *Plan* Operations	Notes
0)	**Achieve Overall Goal**	
	Plan 0: 1 then 2	
	1. Do Task 1	
	2. Do Task 2	
1)	**Do Task 1**	
	Plan 1: In any order: 1.1, 1.2, 1.3	
	1.1. Do Subtask 1	
	1.2. Do Subtask 2	
	1.3. Do Subtask 3	
2)	**Do Task 2**	
	Plan 2: Choose one of: 2.1, 2.2, 2.3	
	2.1. Do Subtask 4	Requires equipment
	2.2. Do Subtask 5	
	2.3. Do Subtask 6	

Figure 1: Graphical and tabular formats of HTA for the same simple task

Applications

The technique was introduced by Annett and Duncan (1967) to solve the industrial training problem of deciding at what level of detail a task should be represented, and since then has been used for many purposes (Table 1). It is commonly used as a preliminary tool for understanding tasks, often with the aid of the many extensions that have been developed (Stanton, 2006).

Purposes	**References**
Team/job-design Information-requirements specification Documentation Human resource management	Shepherd, 2001
Training	Shepherd, 1985
Human error prediction	Stanton *et al.*, 2009; Stanton and Stevenage, 2010
Interface design	Wilson and Rajan, 1995
Measuring team skills	Annett, Cunningham and Mathias-Jones, 2000
Situational awareness measurement	Salmon *et al.*, 2009

Table 1: Example purposes for HTA

Domain	**References**
Aviation	Stanton *et al.*, 2009
Chemical	Hellier, Edworthy & Lee, 2001; Kuselman *et al.*, 2013
Defence	Jenkins *et al.*, 2006
Medical	Chung *et al.*, 2003
Production & Maintenance	Shepherd, 2001
Office	Sebillotte, 1988

Table 2: Example domains for HTA

Many individual examples of HTA have been published, often to demonstrate its use with a new combination of extensions or in a novel domain (Tables 1 & 2). There seems to be little description in the literature of its normal use by human factors/ergonomics (HFE) practitioners.

Criticism

HTA has been criticised primarily for the skill requirement to use it properly (Kirwan & Ainsworth, 1992:118), and its lack of representation of physical objects, propagation effects, causal understanding, and social-organizational knowledge (Miller & Vicente, 2001). Stanton (2006) suggests that extensions correct these, however they only apply to the tabular format. The graphical representation has remained largely unchanged since the introduction of HTA nearly half a century ago.

Learning

Understanding how operator skill is learnt is critical. No studies on reading HTA were found, however Adams *et al.* (2013) presents skills involved in HTA creation based on empirical studies with novices and practitioners. This expands on the similar work looking at how novices create HTAs (Patrick *et al.*, 2000; Felipe *et al.*, 2010). All found that novices had difficulty understanding the hierarchy, and would initially describe tasks in a list format or a flow-chart-like algorithm. Logical decomposition & equivalence - that a task is made up of (and is the same as) its subtasks - were particularly poorly understood.

Understanding the decompositional nature (and to some extent logical equivalence) of the hierarchy may not be as difficult as these studies would suggest. Sebillotte (1988) suggests that when task description is framed as repeatedly asking ‘how?’, people familiar with a task will naturally decompose a task into subtasks hierarchically. Given this, through training and improved notation, it may be possible to help schematize hierarchical structure and decomposition.

Plan Developments

Alongside task-specific changes (e.g. Ormerod & Shepherd, 2004; Stanton & Stevenage, 1998), the plan, another single-element item in a table, has received attention. Shorthand exists for common task sequencing (Table 3).

Meaning	**Plan type**	**Shorthand**
Do in order	Sequential plan	1 > 2
Do in any order	Non-sequential plan	1 / 2
Choose one of...	Selection	1 : 2
Do at the same time	Parallel	1 + 2
Do if condition X is met, otherwise...	Contingent	X? Y > 1 N > 2

Table 3: Plan templates (Stanton, 2006)

Graphical Notations

Graphical vs Non-graphical

Compared to linear information (text), diagrams allow viewers to find information more efficiently because they allow for very fast, easy, perceptual inferences, improving pattern recognition and search strategy (Larkin & Simon, 1987). This reduces cognitive overhead, minimising the effort required by working memory, which appears to be a limiting factor, particularly for complex tasks (Lohse, 1997).

Graphical depictions are not inherently superior to textual information: effects depend on the cognitive fit between user skill task type, and representation type (Petre, 1995; Vessey, 1991; Vessey & Galletta, 1991). Morgan & Michaelson’s (2012) empirical comparison between tabular and graphical presentations for multiple task types suggested there was little difference in performance between them, but graphical presentations were subjectively rated higher on multiple scales.

Subjective ratings should not be dismissed: the more attractive an interface, the higher the expected ease of use (Kurosu & Kashimura, 1995; Tractinsky, 1997).

Graphical vs Graphical

Despite assumptions that notational details are trivial, they are actually very important in conceptual modelling (Hitchman, 2002). Graphical depictions vary widely in the quality of notation that they employ. Informationally equivalent graphs[1] are not necessarily computationally equivalent[2]; representational advantages can offset user familiarity (Peebles & Cheng, 2003); and perceptual complexity, not only problem complexity, affects problem-solving ability for a fault location task (Brooke & Duncan, 1981).

What makes a graphical notation 'good'?

Implicit so far has been the concept of a 'good' notation. Moody (2009) recommends cognitive effectiveness as a metric: how well a notation enables accuracy, ease and speed of use. This, importantly, allows for falsifiable hypotheses and thus improvement claims can be empirically tested.

Petre (1995) stresses the importance of secondary notation - non-formalised notation in a diagram that provides cues that help (or hinder) the viewer (e.g. spatially clustering functionally-related components). This may or may not be intended by the diagram designer.

Moody (2009:774) builds on this with principles based on research in "communication, semiotics,... visual perception,… cognitive psychology, HCI... diagrammatic reasoning". Principles are explained as they are applied.

In the traditional notation (TN), the plan (central to HTA (Shepherd, 2001)), is text only, missing the opportunity for computational offloading.

[1] "information in the one is also inferable from the other" (Larkin and Simon, 1987:67)

[2] "informationally equivalent and... any inference that can be drawn easily and quickly from the information given explicitly in the one can also... from the information given explicitly in the other" (Larkin and Simon, 1987:67)

2. Aims and objectives

The aim of this project is to determine which elements of the graphical representation of HTA would most benefit from improvement, make changes based on practitioner usage and relevant literature on graphical notation and perception, and iteratively test changes against the original design. The objectives follow:

- Contextualisation - prioritise areas of development, determine who uses HTA, how it is used, and what issues are experienced by HFE practitioners.
- Principles - suggest revisions to the graphical notation of HTAs based on Moody's (2009) prescriptive theory of diagrammatic notation.
- Validate and refine these revisions.
- Evaluate the revisions:
    - Empirically compare cognitive effectiveness for the revised and traditional notations.
    - Get rich feedback from expert users.

General Approach to Methodology

The project took a mixed methods approach to determining problems and evaluating potential solutions, incorporating both rich and quantitative data. Multiple studies evaluating and revising graphical notations (using a variety of methods) influenced the approach (Table 4).

Reference	Graphical language	Framework	Methods
Kutar, Britton & Jones, 1997	CSP	Cognitive Dimensions (Green, 1989; Green & Petre, 1996)	Questionnaire to assess notation intuitiveness by untrained users. - Multiple choice questions on which, of 5 symbols, best represented a given word/phrase - Independent groups design
Moody & Hillegersberg, 2008	UML	Nascent version of Physics of Notations (Moody, 2009)	Systematic application of principles in framework
Figl *et al.*, 2013	Compared EPC, UML, YAWL, BPMN	Physics of Notations	- Changes based on principles - Measured model comprehension, perceived cognitive load and time taken - Untrained student participants - Independent groups design: same task and layout, different notation

Table 4: Different methods used to evaluate and revise graphical notations

There is insufficient literature is available on which to prioritise many possible notational changes, necessitating the initial contextualisation study. Additionally, although “expert consensus is *not* a valid way to choose graphical conventions” (Moody & Hillegersberg, 2008:3), user participation tends to improve acceptance of the end result (Gyi *et al.*, 2015). Expert users are unusual for this study, as ergonomists should be more aware of requirements for good design (although Gillan *et al.* (1998) notes that this is not necessarily true).

Loughborough University Ethical Advisory Committee approved the project.

3. Study 1: Contextualisation

Participants

Sampling involved requesting participation on ergonomics and human factors groups on LinkedIn, as well as convenience sampling of practitioners visiting Loughborough University or at conferences the author attended.

Membership numbers of LinkedIn groups:

- Chartered Institute of Ergonomics and Human Factors - 4110
- Human Factors - 13776
- Human Factors and Ergonomics Society (HFES) - 12902
- Ergonomics - 4530

(LinkedIn, 2017a-d)

It is unknown how many of these are active practitioners that use HTA.

22 participants (14 male, 8 female) from many domains responded, including oil & gas, defence, aerospace, maritime, rail, occupational health, patient safety, medical device design, control room design, food safety and academic research.

Method

Participants were asked 12 questions relating to their use of HTA (Appendix C) via e-mail, or in person. This was analysed by recursive extraction, similar to inductive content analysis (Elo & Kyngäs, 2008).

Results

Not all participants use the graphical format, but most do to some extent.

Use of graphical depiction:

- Building/improving mental model of task
- Good for high-level overview (task structure/hierarchy)
- Understanding complexity/detail
- Understandable for non-HFE
- HF team communication
- Good for small/simple task representations
- HTA first pass
- Scanning for patterns/points of interest
- Construction, editing, reviewing, reporting, teaching HTA concepts
- Created and edited by HFE practitioners and non-HFE task experts
- Presented to frontline staff, artists, managers, task-specialists

Desirable features

- Visualised Plans
  - Task sequencing
  - Decision points
- Representation of other model types (flowchart, PCA…)
- Subtask in/output
  - Information-flow between people/machines
  - Dependencies
  - Context and risk information (affecting and affected by tasks)
  - Inhibiting conditions and potential solutions

- Subtask-specific information
  - Difficulty
  - Frequency
  - Perception-Cognition-Action information
  - Importance (safety, performance etc)
    - Levels of criticality
    - Binary highlighter (for safety critical steps)
  - People/roles involved
  - Better distinction between branches and between tasks
  - Editing completion status
- Tools for managing with/teaching complexity
- Explicit links to different pages of the output.

Conclusions - design goals

- cognitive effectiveness (accuracy, ease, speed)
  - adherence to Moody (2009) principles
- familiarity for existing practitioners
- potential for further extensions
- improved learning potential
- appropriate for varied experience levels
- internal consistency
- ability to incorporate flow charts/similar information
- changes that can be used independently of each other

4. Notation Analysis and Redesign

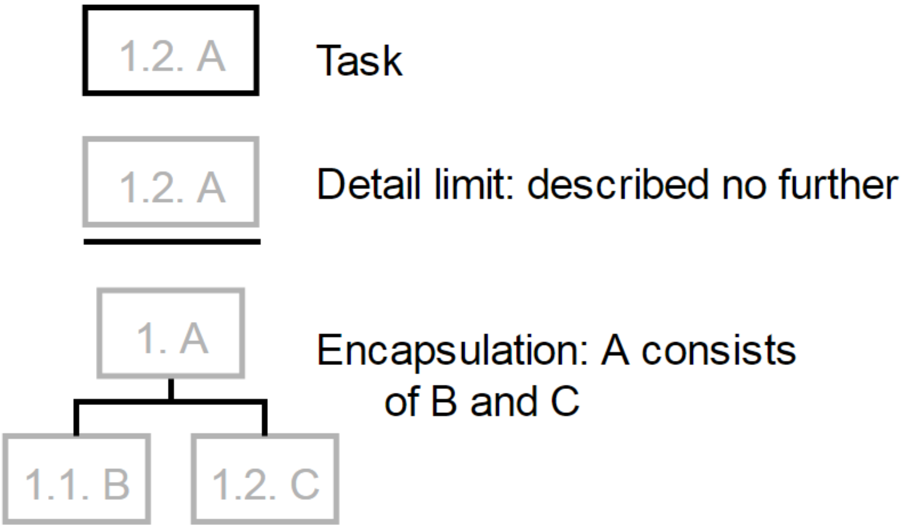


Figure 2: Existing elements of HTA notation

Moody (2009) provides a number of prescriptive principles for analysing and designing graphical notations, and acknowledges trade-offs may need to be made (Figure 3). Except where stated, the following sections are based on this source.

	Semiotic Clarity	Perceptual Discriminability	Semantic Transparency	Complexity Management	Cognitive Integration	Visual Expressiveness	Dual Coding	Graphic Economy	Cognitive Fit
Semiotic Clarity								±	
Perceptual Discriminability						+			+
Semantic Transparency		+							±
Complexity Management								−	+
Cognitive Integration	−			+				−	
Visual Expressiveness		+						+	±
Dual Coding									+
Graphic Economy		+		+		−			+
Cognitive Fit									

Figure 3: Graphical notation principles and their tradeoffs. Symbols indicate how the row element affects the column element. (Moody, 2009)

Semiotic Clarity

The mapping of semantic to syntactic constructs.

There is a large symbol deficit: despite all the semantic constructs of the HTA (and its extensions) - sequencing of tasks, decision-making, task-specific information etc - only 3 are typically shown graphically (Figure 2). This deficit suggests poor representational fidelity (Burton-Jones *et al.*, 2009).

Boxes around plans contributes to symbol excess, as the text of the 'plan' does not map to a single semantic construct, but a box suggests this.

Proposed additional syntactic constructs, based on plan templates (Table 3) are: sequential, non-sequential, either/or, concurrent, and 'may not occur', which suggests a decision is made. Plans will be shown as simple text.

Perceptual Discriminability

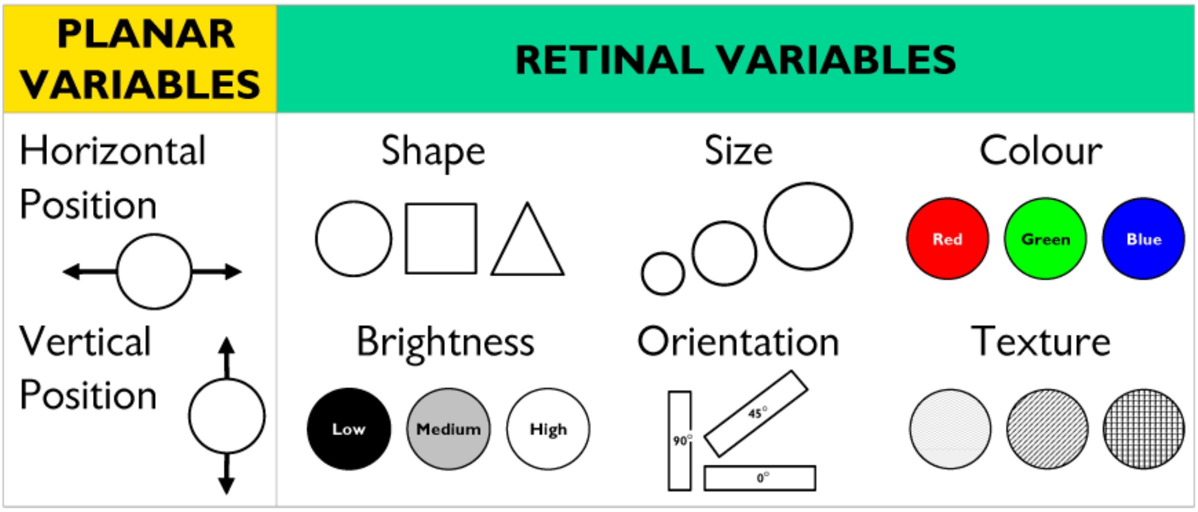


Figure 4: visual variables (Moody, 2009)

Graphical elements can vary in a number of dimensions (Figure 4). Elements that vary more in one dimension are easier to distinguish. Element types that have a unique value on at least one axis (e.g. the only type in red) can enable 'perceptual popout', where the existence of elements is obvious without conscious effort.

The 3 existing notations have different shapes, but no significant variations in other variables. The lines tend to all be rectilinear, which makes differentiation effortful.

Vertical position is used to represent level of abstraction, and horizontal position informally indicates ‘do sequentially left-right unless the plan states otherwise’ (vertical and horizontal are sometimes transposed).

Visual variables have different capacities for representing data: shape encodes nominal, while size can represent ratio data (Carpendale, 2003). Colour is the most cognitively effective variable (differences are detected 3 times faster than shape), however this information should be redundantly coded another way to remain robust when e.g. viewed by a colourblind person or printed monochromatically.

Shape and colour may lend themselves well to domain- and task-specific information e.g. Perception-Cognition-Action coding (Rauterberg, 1999).

Semantic Transparency

Intuitive understandability of symbols. Symbols can be:

- ***‘Transparent’** - meaning can be inferred from appearance (e.g. stick figure → person).*
- ***‘Translucent’** - appearance suggest their meaning (act as mnemonic) but require initial explanation.*
- ***‘Opaque’** - relationship between appearance and meaning is arbitrary.*
- ***‘Perverse’** - appearance suggests a meaning different than intended.*

Boxes as a task are abstract, and so opaque. Detail limit is more translucent, as it is indicative of a barrier.

The branching lines of a hierarchy are commonly used and show a connection, but do not well convey the concept of encapsulation - logical decomposition and equivalence. As well as the concept of ‘containment’, hierarchical/tree depictions can also depict a ‘top-down’ expansion from a starting point (phylogenetic/evolutionary trees (Ciccarelli *et al.*, 2006:1284)) or ‘bottom-up’ prerequisites/routes for a single endpoint (Gagne’s (1968:4) Hierarchy of Learning). The appropriate type of decomposition is not always well understood (Patrick *et al.*, 2000).

Hierarchy syntax should better convey the fact that subtasks are ‘the same’ as the task they make up. Based on theories of perceptual proximity (Wickens & Carswell, 1995), objects can be visually integrated using contiguity, contour, and spatial integration.

Complexity Management

Complexity is noted as an intractable problem in notation design, as diagrams tend to scale poorly. HTA is already built around the two main techniques for managing complexity: hierarchical decomposition and modularisation. The existing mechanisms could be better represented (see above).

Cognitive Integration

If multiple diagram types are used to represent the same system, the ability to refer to others is useful. Space for additional information would be useful, and can be combined with the proposed approaches to hierarchy. Further revisions are outside this project’s focus.

Dual Coding

Suggests that important information should be graphically *and* textually represented; these are not expected to visually conflict. This applies to the plan.

Graphic Economy

Notations should be no more complex than necessary. Complexity must be balanced with expressiveness. A line from a task to every subtask to indicate decomposition may be excessive.

Cognitive Fit

Different representations may be needed for different use cases and different uses. Novices benefit from “more discriminable symbols… reduced complexity... more mnemonic conventions... explanatory text... simplified visual vocabularies” (Moody, 2009:773). Other than emphasizing these principles, different syntactic constructs should exist independently of each other, so practitioners can determine an appropriate level of complexity for their audience.

Representational medium also affects notations: some visual variables are much harder to depict hand-drawn than on a computer.

As this project addresses the fundamentals of HTA, ability to be represent these in any medium is paramount.

Conclusion

Two preliminary notation revisions were made based on the principles to compare with each other and the traditional format (Figure 5).

5. Study 2: Validation

Aims

To assess semiotic clarity, perceptual discriminability and discover potential areas of improvement, a comparison and analysis of notations was needed. To mitigate systematic bias for or against one notation, participants were chosen who had not seen any before.

In this early stage, rich feedback was considered more important than strict internal validity, and as questions would likely be asked by the participant to the researcher, a semi-structured interview format was decided upon.

Participants

N = 6 (5 male, 1 female). Age range: 24-60. Professions: chiropractor, game developer/technical artist, careers development consultant, housing manager, head of PE, science teacher.

Design samples

The format of the samples was condensed to show the notation without extra information, in order to focus the comparisons. As such, the plan and hierarchical numbering (e.g. 2.2.1) was removed, and the text for each task was replaced with a letter, to mark where labelling would exist. A minimal structure was chosen to show each of the notational constructs (Figure 5).

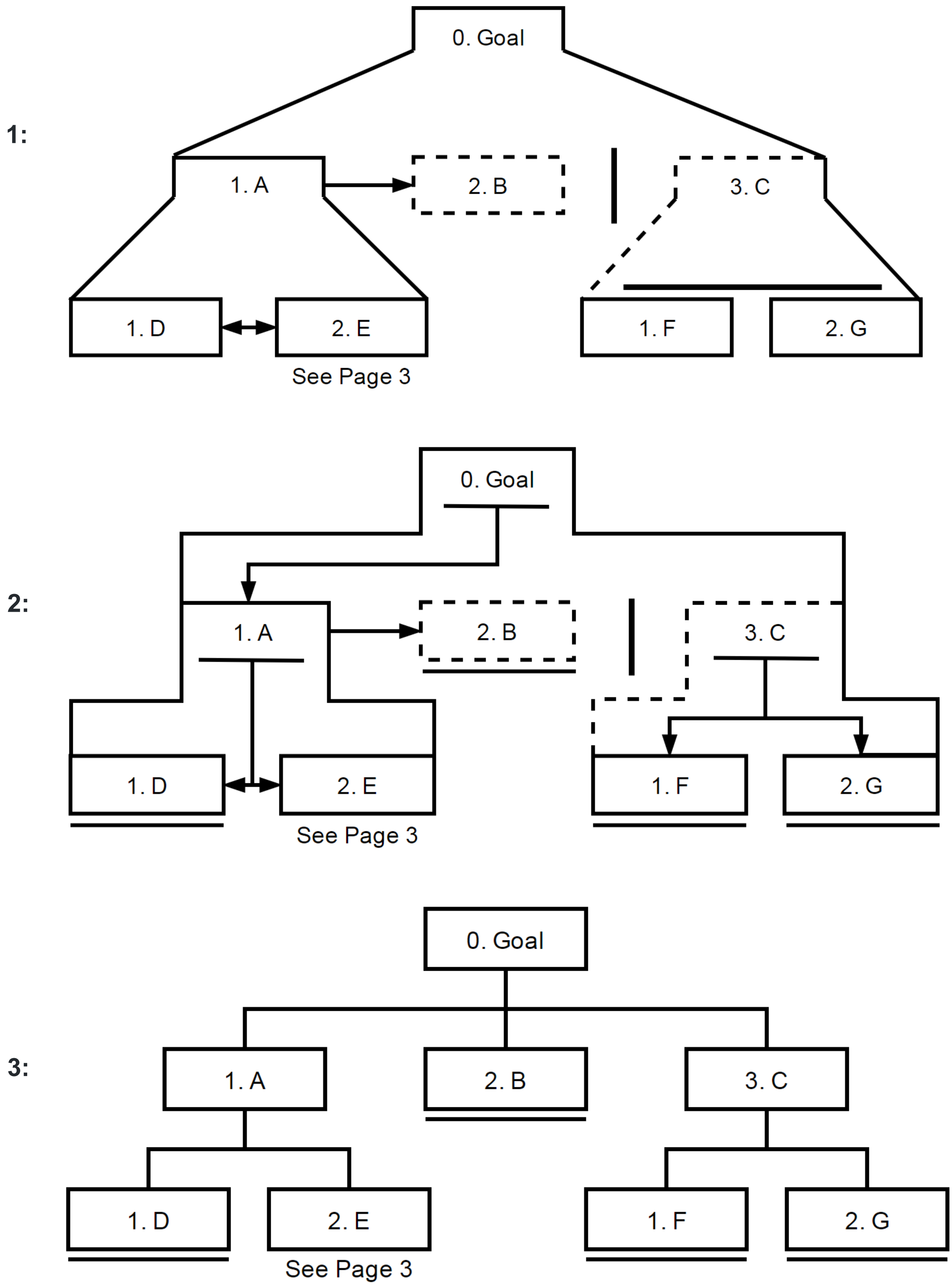


Figure 5: The initial 3 designs shown to participants: (1) implicit routing, (2) explicit routing, (3) no routing (traditional).

Procedure

The 6 different variations of ordering (123,132, 213...) was randomised for each participant using a die to minimise systematic order effects. The participant was shown the first representation, and asked to explain what they thought it showed, given that it represented a task. They were shown the next representations and asked if they thought they showed anything different, or changed their opinion on what was shown. They were then progressively informed on what tasks were supposed to represent to see how their understanding changed. They were asked their opinion of each graphical element, with comparisons between notations and if they would change anything. Any design-related information was written down for later analysis. A 4th design incorporated revisions from each participant. This evolving design was shown after participants had seen the others. They were asked which of the representations they preferred (and why).

4:

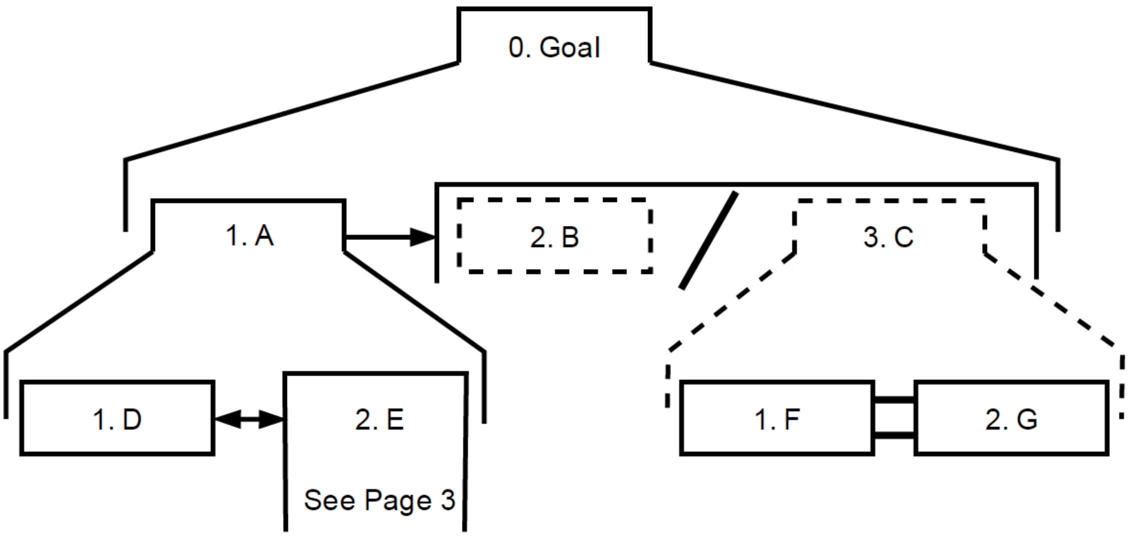


Figure 6: Notation 4 - emerged from participant feedback

Results

Hierarchy/Encapsulation/Grouping

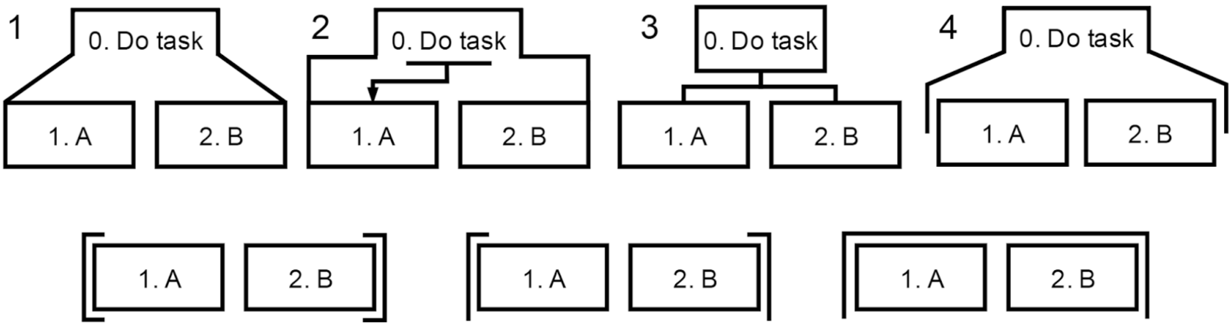


Figure 7: Encapsulation and grouping styles side-by-side

- Treated by some as top-down/bottom-up links for notations 2, 3 and less for 1. The term ‘goal’ strengthened this association. 3 suggested ‘classification’.
- One, already familiar with hierarchical depictions in the style, found 3 easier
- The contiguous task contours for 1 and 2 led eyes around the outside, visually connecting semantically unconnected tasks.
- Diagonal lines preferred to rectilinear: stood out more from tasks. Rectilinear lines all seemed similarly important: it was hard to know where to look.
- The ‘umbrella effect’/‘magnification’ made 4 semantically transparent as decomposition to most. The space at the sides stopped leading eyes distractingly and implied subtasks are ‘done within’ the task. Considered best.
- Grouping (B-C) was considered helpful. 3 grouping types were tried with minimal difference in preference. The third suggested the contents were treated as a single task. ‘Visual brackets’ lower reliance on operator precedence rules.

Sequencing

Figure 8: Sequencing notations

- One-sided arrow: semantically transparent to most as sequential. One thought it could be a side-effect.
- Two-sided arrow: less semantically transparent, but mnemonic when explained as ‘any order’- “the penny drops”.
- Horizontal bar: suggested that an outcome occurred from the combination of the elements under it; this and the two-sided arrow were seen as equally possible options for ‘any order’ or concurrent.
- Parallel lines: slightly preferred to horizontal bar for mnemonic ‘do in parallel’
- Vertical bar: seen as a barrier/separator between B and C. Semantically translucent that this meant ‘or’.
- Diagonal slash: more visually distinctive - stands out from task and encapsulation lines. Mnemonic/transparent as commonly used to denote alternatives in writing (Waddingham, 2014)

Possible occurrence

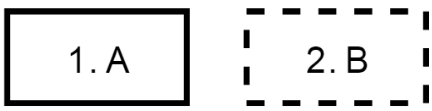


Figure 9: Definite and possible occurrences

- Dashed box suggests something one “may need to do, maybe not crucial, something to consider”. Semantically transparent to most.
- The half-dashed, half-solid surround of C was confusing (Figure 5); 4’s encapsulation avoided this problem.

Detail limit

Figure 10: Detail limit reached and not reached for notations 3 and 4

- An open/closed box was a well-liked visual metaphor that existed across tasks, encapsulation and grouping. It was also visually distinctive.
- The underline was also understood as the end of a section, showing the outcome of the task, or as a requisite for task completion.

General Comments

- 1 avoids the excess lines of 2 and 3.
- 2 looked harder and more confusing than others, with more irrelevant information.
- 3 thought more organised than 1 or 2 by one, but another felt that its not showing B OR C 'ruled it out'.
- 4 was visually the easiest/clearest, if slightly complicated. One understood every feature without any hinting (compared to a maximum of a couple of features in other notations).
- Task as a box is opaque but easy to learn.
- All preferred visual indication of the plan to just an accompanying written plan.
- Multiple suggestions were made:
  - A dashed connection or lack of notation should be considered 'any order'
  - Sequential tasks should have y-axis position show ordering
  - Parallel tasks should be vertically stacked
  - Parallel tasks should have arrows facing towards each other: →←
  - Tasks described elsewhere could be applied as subroutines
  - Card suits could be used as additional differentiators

Conclusion

There is a revised notation (Figures 6 & 11) that appears to improve upon the traditional format, which can be tested.

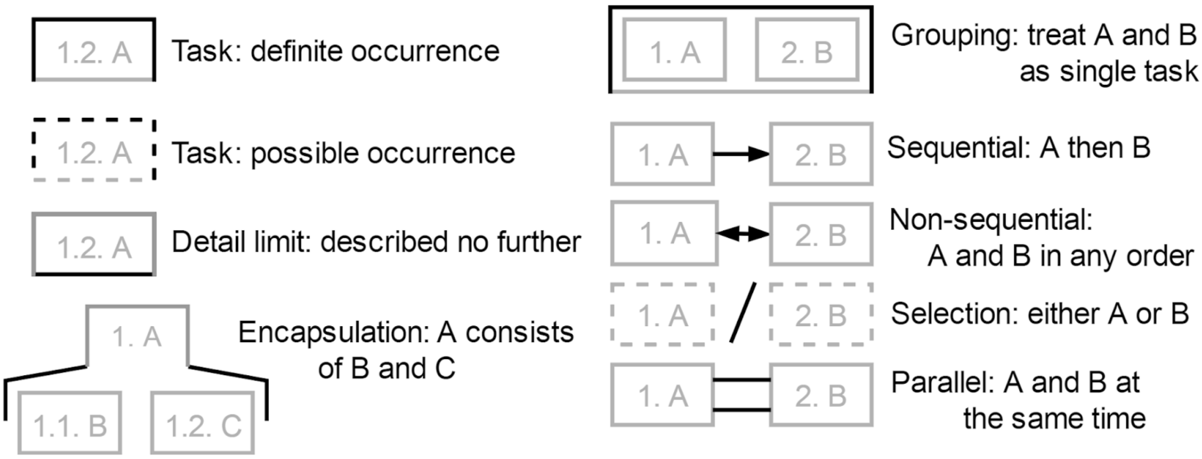


Figure 11: Revised notation

6. Study 3: Empirical evaluation

Methodology

To determine the change of notation as the cause of any difference in results, this study would use an experimental approach. Different settings could have been used, which would have emphasized different aspects of validity. A more naturalistic setting that focused on replicating a real-world use of HTA would have high ecological validity, but may risk introducing potentially confounding results or restricting generalizability.

As a result, the intended approach should be based on the core components of HTA, requiring familiarity with only the basics of HTA and none of its extensions. A laboratory setting should increase the internal validity of the study by minimising extraneous factors, with a small tradeoff in terms of ecological validity.

A paper questionnaire was used, which included a tutorial. Here, the students had just finished a lecture on the use of HTA, fulfilling this requirement.

Sample selection

Given that the sample size would not be large, a heterogeneous sample could provide too many extraneous factors, making definitive conclusions difficult. A homogeneous sample was preferred. Particularly important was for participants to have a similar level of skill/training in using HTA.

A convenient group was the class of 2nd year Ergonomics students at Loughborough University. The time of the experiment was immediately following their second lecture on HTA.

Participants

N = 29 (16 male, 13 female), age range: 18-29.

Task creation

Given knowledge of the likely participants, a task should be chosen so as to be familiar and demonstrating a number of the notational constructs created. The task of making scrambled eggs on toast is both familiar to many and involve all the plan sequence types. Diagrams were created in original and revised notations.

Questionnaire design

Two main evaluative approaches were decided upon for the questionnaire:

- Measurements corresponding with accuracy, ease, and speed
- Evaluation of principles aimed at improving this.

A primarily ‘independent groups’ design was chosen, as a new second task would take too much time for the participants to engage with, presenting both logistical difficulties and a likely decrease in response quality (Galesic & Bosnjak, 2009).

Accuracy was determined by the number of questions on the nature of the task that were answered correctly. Each notation type was tested by a question. Questions resembled surface-level comprehension tasks as these assess the interpretability of the notation, which is “the basis for a variety of more specific tasks such as process analysis or redesign” (Figl *et al.*, 2013:321). Ease was rated by participants on visual analogue scales (Maxwell, 1978) presented after each question and one following all questions. Other subjective ratings were taken the same way. Speed was determined by the difference in time when participants began and ended Section A.

The questionnaire was piloted with four people familiar with HTA to ensure the questions were understandable, would yield appropriate responses and that the diagrams were informationally equivalent with respect to the questions. Following their feedback, questions were rephrased and one was replaced.

Procedure

The study was briefly introduced to the class of students, noting that it would provide them an opportunity to practice using HTA. They were provided with participant information sheets and consent forms (Appendices A & B), and it was reiterated that their participation was optional.

Participants were provided with the questionnaire and task depiction corresponding to their notation group, which alternated for students sitting next to each other.

The continuous nature of the rating scales was emphasized, with an example answer shown. Instructions were projected onto a screen at the front of the room for the duration of the study. Participants were asked if they had any questions before starting. Responses to all remaining questions were deferred until all participants had finished section A to ensure participants would receive the same information. After all participants had finished section B, they were asked to swap task representations with their neighbours, so that they could answer the comparative questions in section C. Once all participants had finished, their questionnaire booklets were collected, and they were debriefed.

Ratings were measured, then all responses were transcribed verbatim into SPSS, coded and normalised where appropriate. Answers for Section A were marked by comparing with a master copy. Missing answers were considered incorrect.

Results

Multiple measures violated the assumptions of normality in their distribution and/or had clear outliers. It follows from this that a nonparametric approach should be used (Leech & Onwuegbuzie, 2002). As the efficiencies of the relevant nonparametric tests are close to their parametric counterparts even under conditions of normality (Sawilowsky, 2005), the nonparametric approach was applied to all measures for the sake of consistency. Effect sizes are given as the rank-biserial correlation *r* (Cureton, 1956) due to its ease of calculation and interpretation for both independent and dependent samples (Wendt, 1972; Kerby, 2014). Results were considered significant when $p < \alpha$, *where* $\alpha = .05$. SPSS was used to calculate exact *p*-values.

Hypotheses took the following general forms, where *A* and *B* indicate a notation condition:

Mann-Whitney U (Nachar, 2008):

H_0: responses for *A* and *B* have the same distribution

$H_{1\ (1\text{-tailed})}$: responses for *A* will be stochastically greater than for *B*

$H_{1\ (2\text{-tailed})}$: responses for *A* and *B* will differ in distribution

Wilcoxon signed-rank (McDonald, 2009)

H_0: median difference between observation pairs is zero

H_1: median difference between observation pairs is non-zero

Accuracy

Participants were expected to perform better on the modified representation, so tests were one-tailed. Accuracy was determined by the number of questions answered correctly.

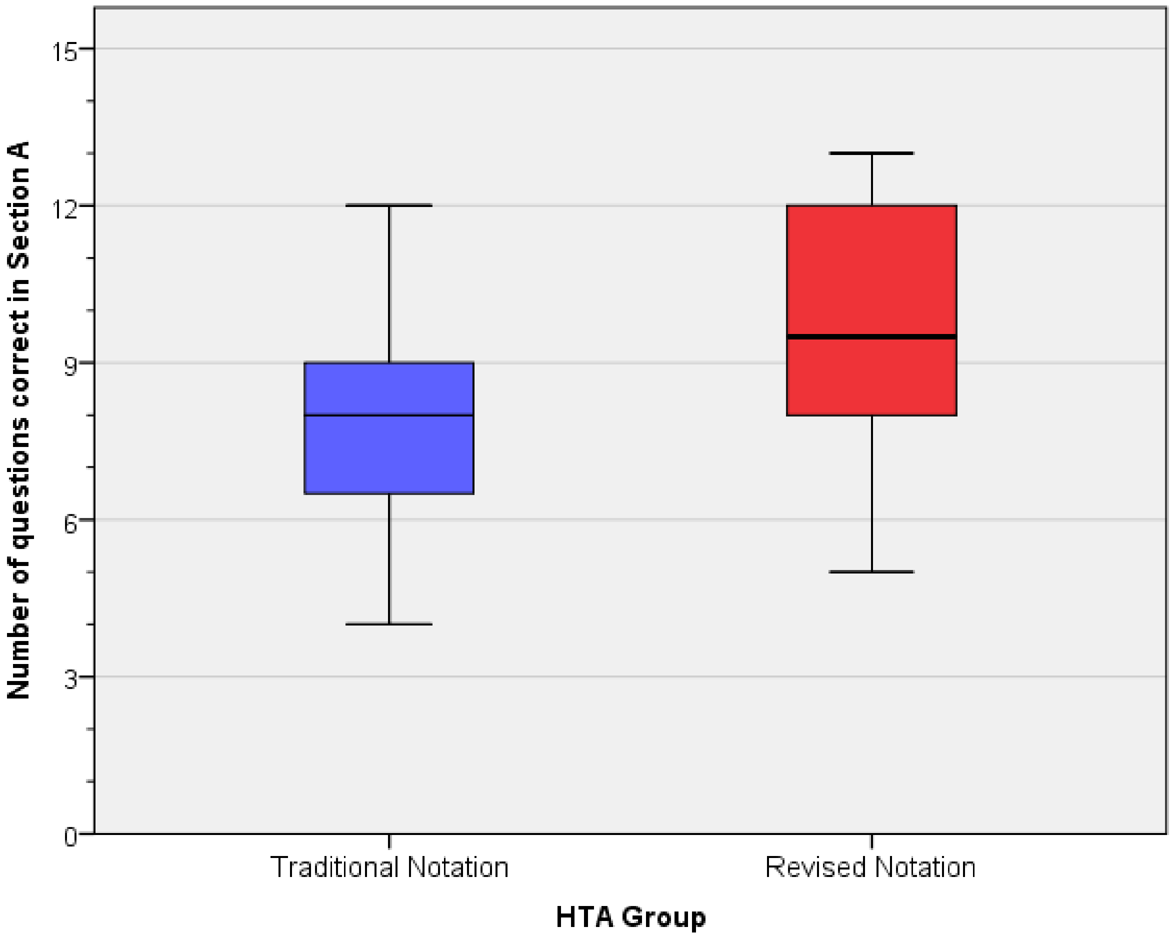


Figure 12: Boxplot comparing numbers of correct responses for TN and RN

The number of questions answered correctly (QAC) in Section A provides an overall assessment of the accuracy attained in each group. The median QAC for RN is 1.5 more questions correct than TN. The inter-quartile range of RN is larger than TN. A one-tailed Mann-Whitney U test demonstrated a significant difference between the two groups: ($n_{TN}=15$, $n_{RN}=14$, $U= 56.5$, $p=.016$, $r=.462$)

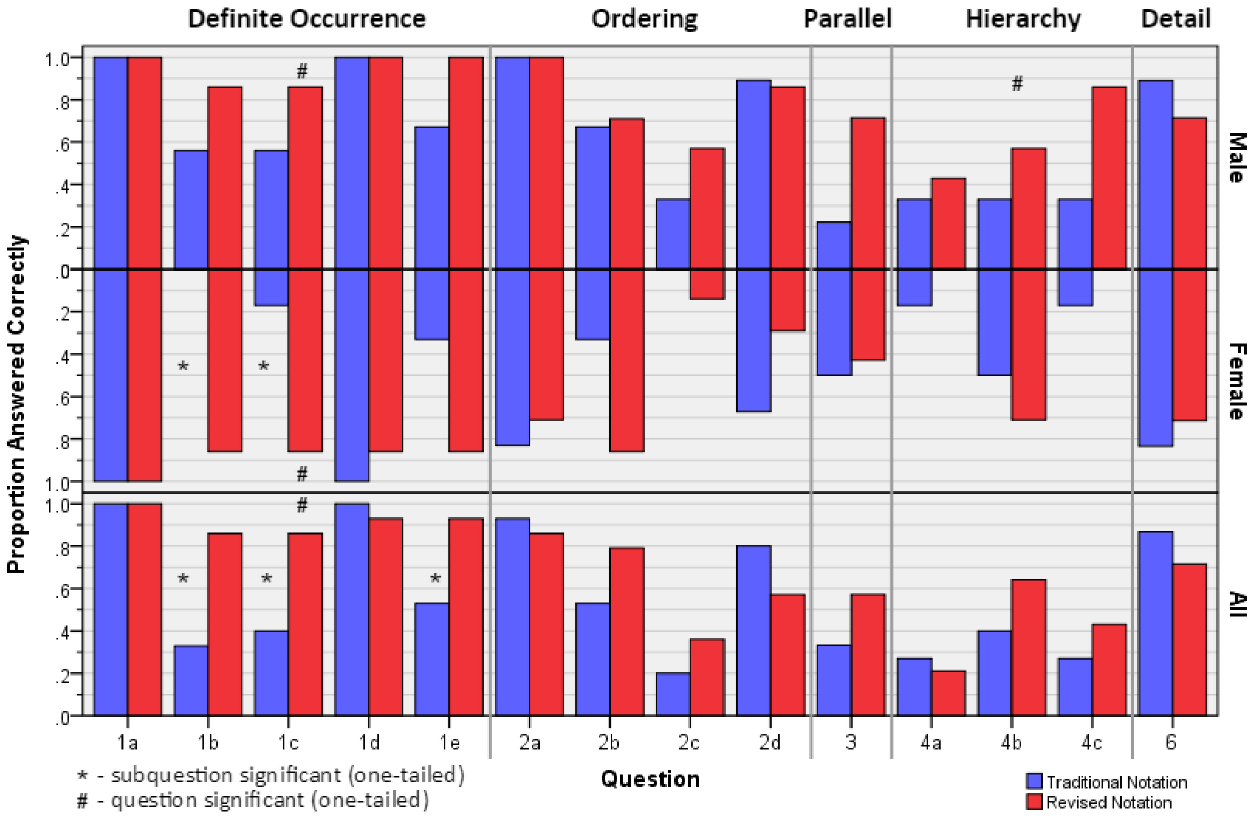


Figure 13: Bar chart comparing proportion of correct answers between notations for each question

The only numbered question with significantly more correct responses across all participants in RN than TN was 1 ($n_{TN}=15$, $n_{RN}=14$, $U=47$, $p=.003$, $r=.552$), determining which tasks would definitely occur. Within this, the subquestions b, c, and e (all for non-definite tasks) were answered significantly more correctly in RN than TN (p=.008, p=.0185, p=.035 respectively)

Male participants were significantly more accurate for questions on hierarchy with RN ($n_{TN}=9$, $n_{RN}=7$, $U=13$, $p=.033$, $r=.587$).

Nobody answered question 5 correctly, so it was excluded from the analysis.

Ease

Directionality of ease was not predicted, so tests are 2-tailed.

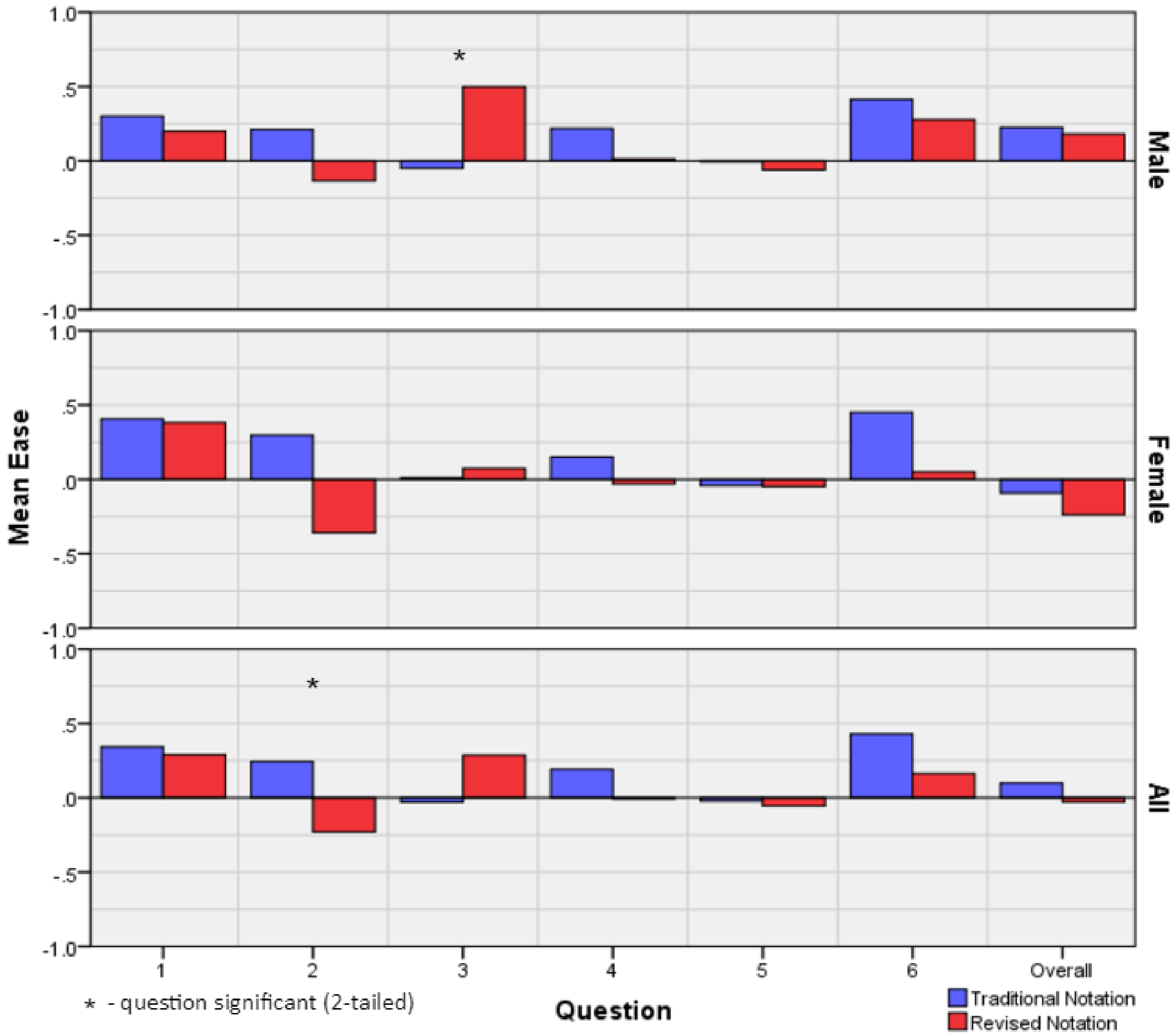


Figure 14: Bar charts comparing ease rating of questions by gender and notation

Except for question 3, all mean ease ratings were lower for RN than TN. For 1, 5 and overall, this was negligible. Question 6 had a larger difference, but it was not significant (n_{TN}=15, n_{RN}=14, U=79, p=.265, r=.248), and neither was question 4 (n_{TN}=15, n_{RN}=14, U=69.5, p=.125, r=.338). Question 2 was considered significantly less easy for RN than TN (n_{TN}=13, n_{RN}=12, U=28.5, p=.006, r=.635).The difference in ease of question 3 for females was negligible, but for males, RN for this question had the highest mean ease rating; this was also the largest difference in means, and was statistically significant ($_{TN}$=9, n_{RN}=7, U=12.5, p=.045, r=.603).

Speed

RN was expected to be faster, so tests are 1-tailed.

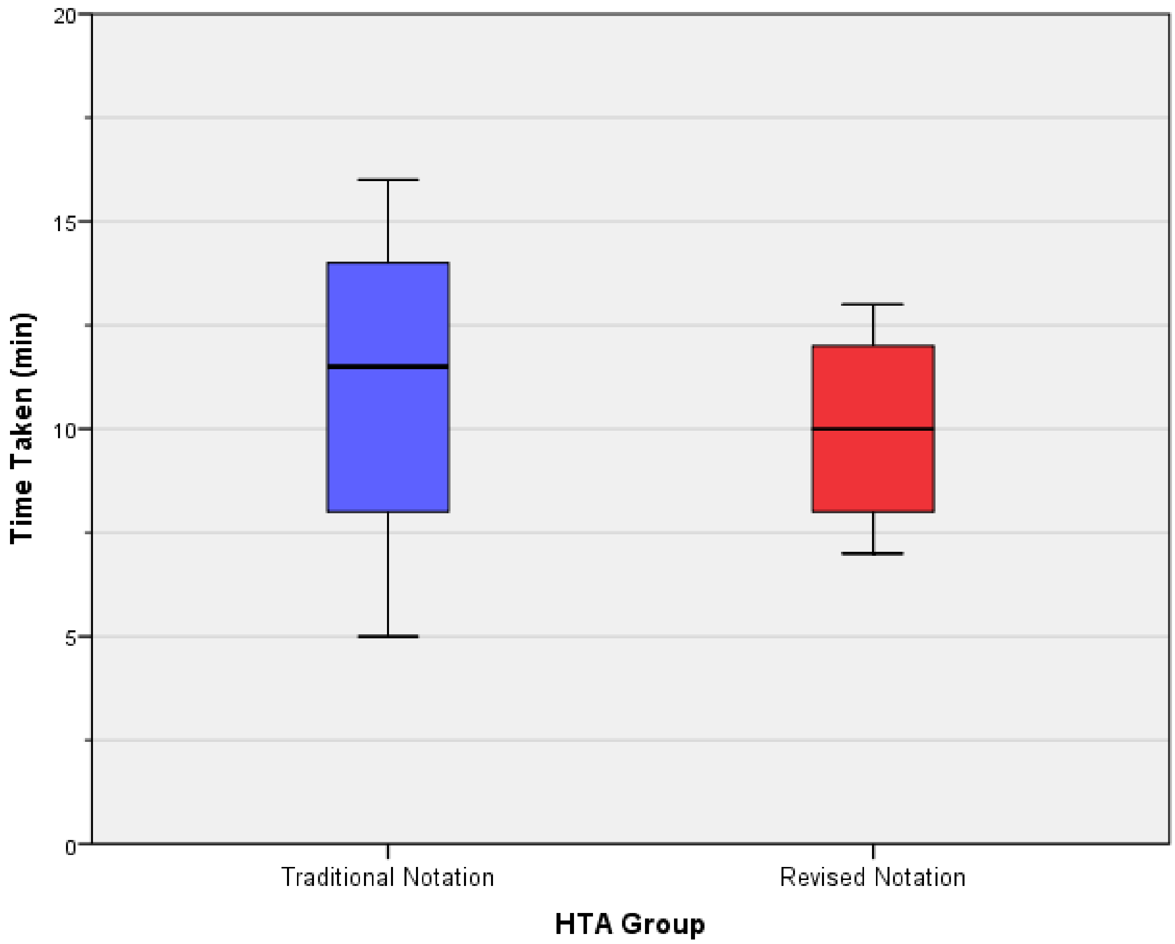


Figure 15: Boxplot of the time taken to complete section A with each notation

Lower median time taken for RN than TN. The range (and IQR) for TN is much larger than RN, possibly because results at the low end were participants who had not fully understood/engaged with the questions. This matched the observation during the experiment that some participants were progressing much faster than expected based on the pilot study. As it is not a useful metric to find out how quickly wrong answers can be put down, this was accounted for by dividing time taken by the number of questions answered correctly across section A. This also provides a more generalisable metric of performance.

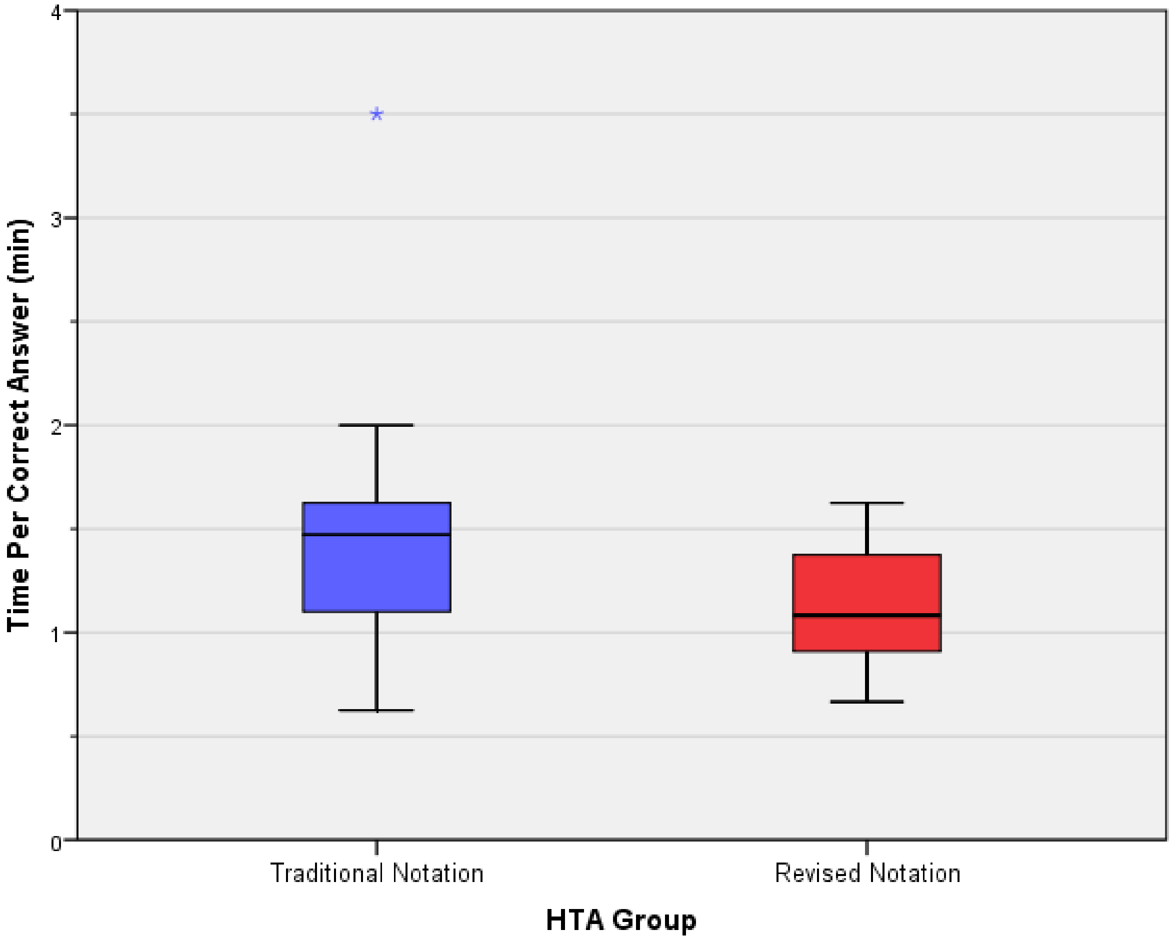


Figure 16: Boxplot showing times taken per correct answer for each notation

The time taken per correct answer (TPCA) appears more homoscedastic between RN and TN than time alone. The median (appropriate considering the presence of outliers) TPCA is notably higher in TN (1.4722 minutes) than RN (1.0833 minutes), taking 35.9% more time for each correctly answered question. A one-tailed Mann-Whitney U test demonstrated $TPCA_{RN}$ was significantly smaller than $TPCA_{TN}$ ($n_{TN}=14$, $n_{RN}=13$ $U=52.5$, $p=.031$, $r=.423$).

Other ratings

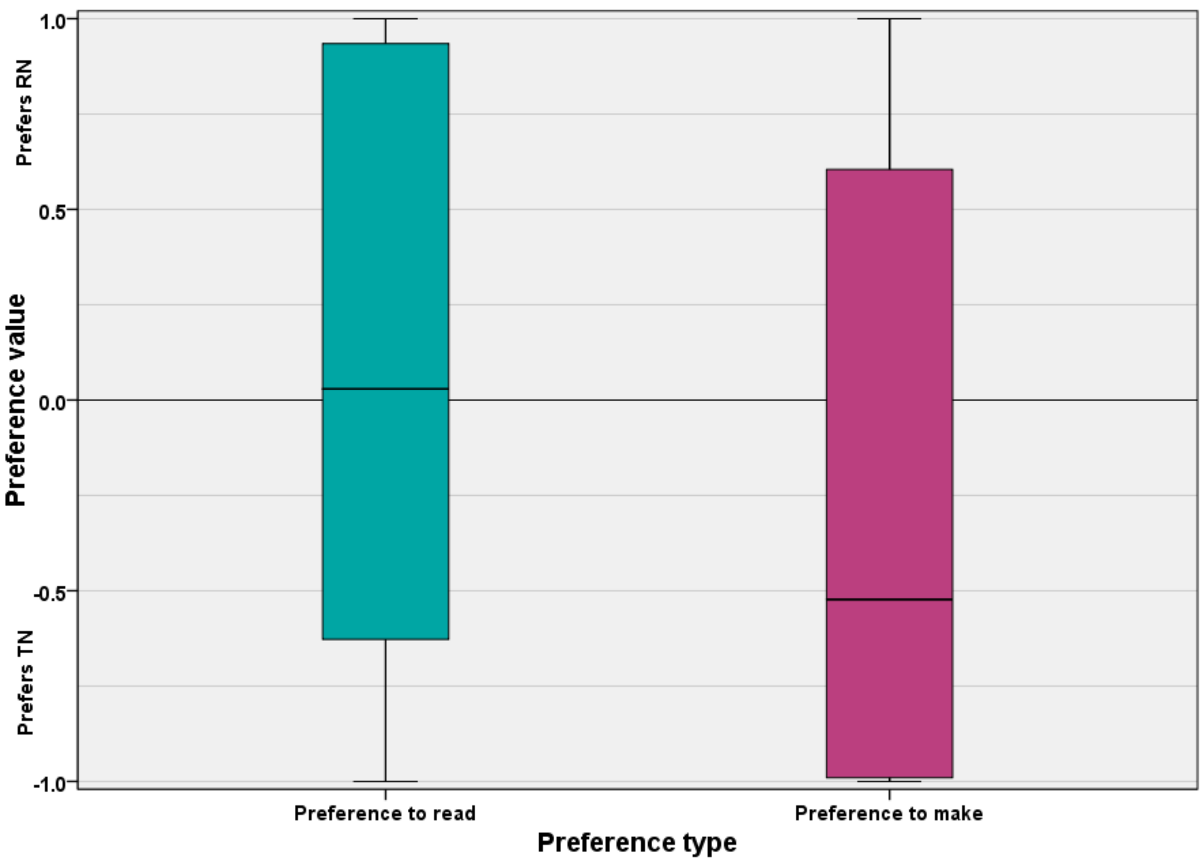


Figure 17: Box plot showing which notations participants would prefer to read and make

All significance tests on preference were binomial.

Read preference spans the full possibility space with a median near zero. RN was not significantly preferred (*p=1.0 mean=7.857, 95% confidence interval: [-14.555, 30.27]*).

The range of write preference was also maximal, but the median is moderately negative. TN was not significantly preferred (*p=.572, mean=-14.661, 95% CI: [-38.56, 9.239]*).

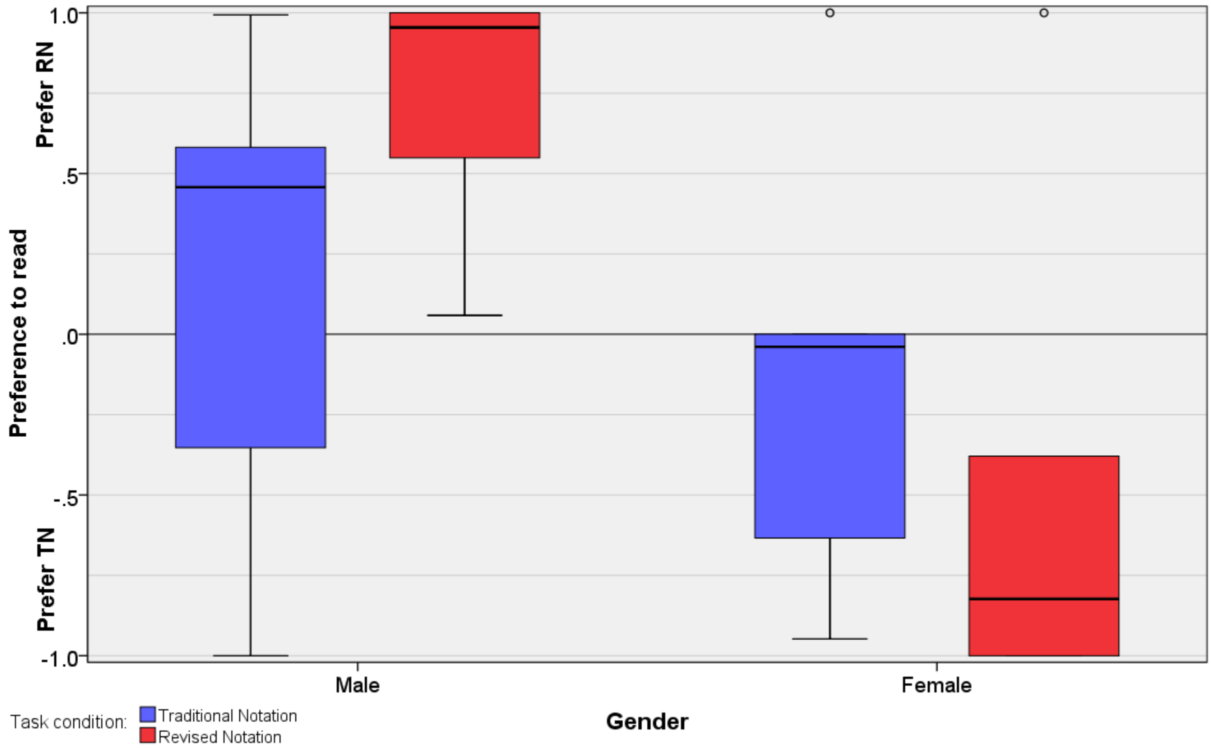


Figure 18: Box plots comparing which notation participants would prefer to read, based on their initial notation (colour) and gender (grouping)

Males significantly prefer RN (n=16, *p=.0385, mean=31.375, 95% CI: [5.276, 57.474]*). Preference was higher if they answered task questions on it first

Females insignificantly prefer TN (n=12, *p=.572, mean=-23.5, 95% CI: [-58.6, 11.6]*).

There was no significant difference overall by students, but significantly more males preferred to read RN if they had used it for their task questions, and the converse was true for females.

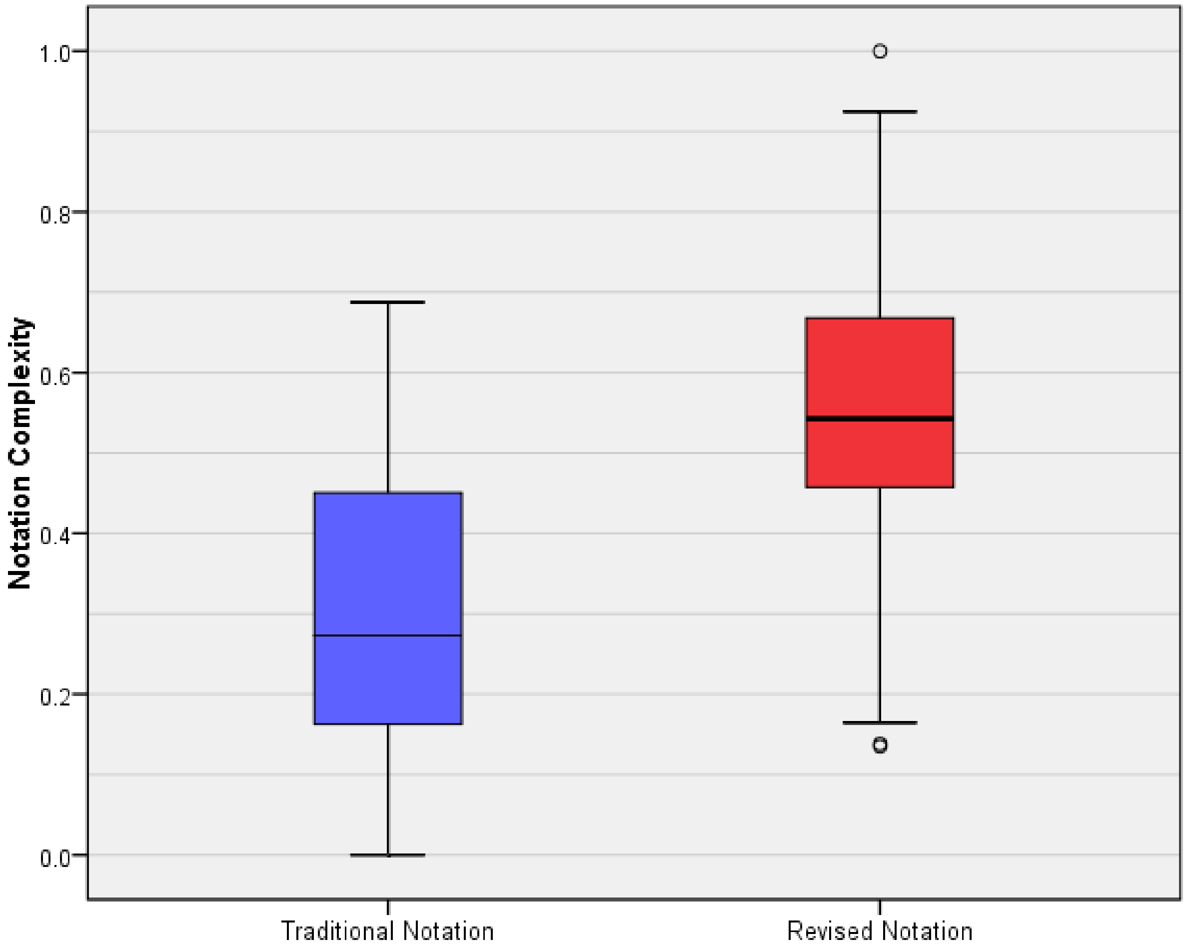


Figure 19: Boxplot comparing ratings of notational complexity for each notation

Directionality was not predicted, so a 2-tailed Wilcoxon signed-rank test was used:

Both diagram complexity (n=28, W=198, p=.023, r=.488) and notation complexity (n=28, W=290, p=.001, r=.714) are significantly different between notations. RN is more complex for both measures.

Conclusions

RN shows some improvements over TN - showing optional tasks improved accuracy for everyone, the new hierarchy did the same, but only for males. Perceived ease was (except for males finding a parallel task) similar or lower for RN. TPCA was lower for RN. RN was preferred to read by males, but not females. RN was considered more notationally and diagrammatically complex.

7. Study 4: Expert evaluation

Method

Following a pilot interview to determine the questions and materials needed, HFE practitioners who used HTA were sent, along with accompanying questions, a number of reference images:

- The original notation and revised notations in their minimal forms (Figures 5 & 6)
- Both versions of the scrambled egg task from the empirical study
- An additional HTA, in the revised notation, of making a wooden dovetail joint, with additional colour categorization by tool (Alexander *et al.*, 2017)

Participants were informed of the aims for the notation, but no details of any results. Their responses were then followed up where additional information might clarify points, and any additional responses were noted. This minimally biased initial results, and allowed a more informed opinion after responses.

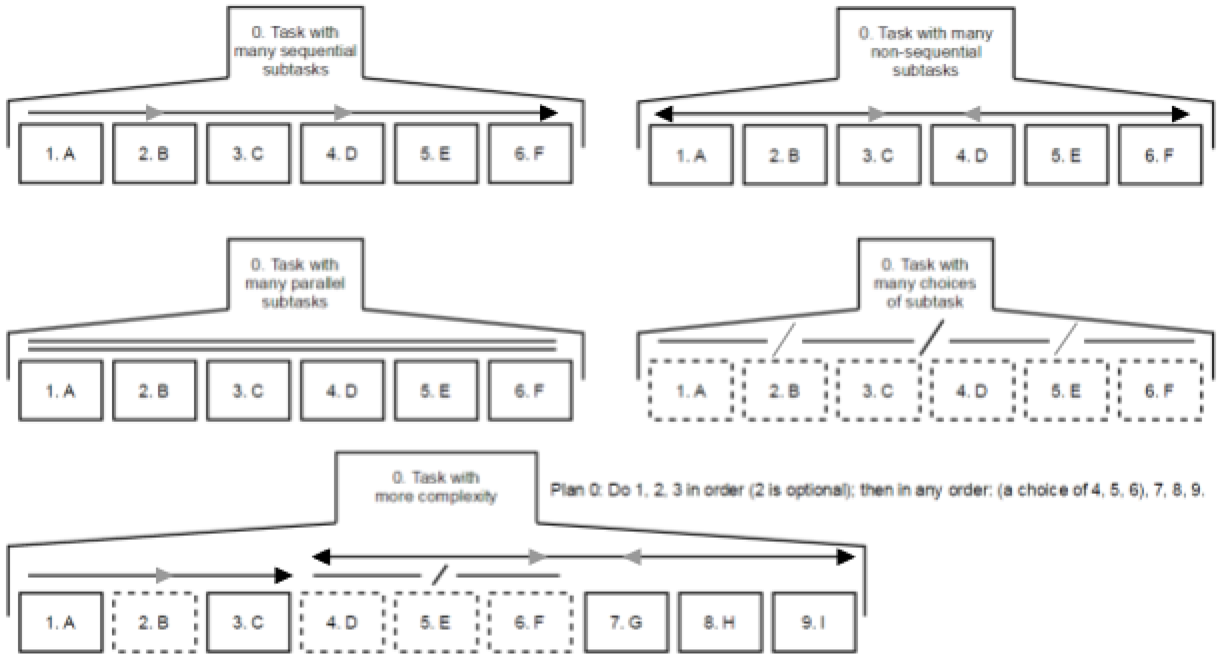


Figure 20: suggested simplifications for sequencing via 'subgroup ordering', with and without repeated notation (grey)

Participants

Participants were sampled in the same way as for the contextualisation study. Participants who used graphical HTA from that study were contacted again.

N = 23 (19 male, 4 female), from a similar range of domains to the contextualization.

Results

Results reflect points/themes that arose. Some are contradictory.

Uses formatting: “direct quote”, ‘paraphrase’, *comment by author*

General feedback

- "There is a need to improve and better formalise the HTA"
- *Some participants would definitely use some or all of the notation*
- *Others would use given certain conditions (addressed later)*
- *Some would not use: happy with existing notation/do not use a graphical format*
- ‘Caution adding too much information/too many visual variables’
- ‘Would benefit from real-life/more complex demonstration of use (for a related domain/purpose)’
- *Improvement for some, no better for others:*
    - "Much easier to follow than a traditional HTA"
    - ‘Draws attention to different structural elements and patterns’
    - ‘No more or less readable (for HFE practitioner)’
    - ‘Redundant options depending on reader preference for graphics/text’

Hierarchy

- ‘More meaningful that a task 'contains' its subtasks’
- ‘Easier to process - strong visual grouping’
- ‘Clearer for novices to see how HTAs are constructed’
- ‘Not needed for HFE experts or when explained to intelligent readers’
- ‘Uncomfortable that tasks are not joined directly to their subtasks’

Plan/sequencing

- Benefits of graphical plan to written plan
  - ‘Faster/easier reference than reading’
  - ‘Simplicity/Consistency’
- Attitude towards specific notation
  - Selection - *most positive response*
  - Sequential - *well liked*
  - Parallel - *liked*
  - Non-seq - *positive, but least so*
  - Described in detail elsewhere - ‘semantically relevant notation’
- ‘Elements are visually distinct’
- ‘Unfamiliar’

Alternative plan notation

- *General preference for this over standard*

Potential tasks

- ‘Very useful - obvious at a glance’
- ‘Clear indication of flexibility/number of decisions needed’
- ‘Does not seem to create any additional cognitive load’

Grouping

- *The addition of grouping was liked*

Complexity

- ‘May be harder/more time consuming to make’
- ‘Notation is more visually complex’
- ‘Diagrammatic complexity is a big problem’
- ‘Does not address diagrammatic complexity’

Non-expert use

- ‘Would need to be explained to non-experts’
- ‘Non-experts may find this harder to interpret’
- ‘Intuitive for non-experts’

Resistance

- ‘Software/time difficulties for creation’
- ‘Requires long-time users to familiarise themselves/retrain’
- ‘Requires full team/department to adopt, overcoming momentum of existing tradition, tools, guidance etc’
- ‘Adoption would be fast/universal if initiated by regulator’

Additional semantics to represent

- ‘Context/dependencies/conditions for tasks and decisions’
- ‘Purpose of task’
- ‘Timeline’

Conclusions

The notation is an improvement in multiple ways and notation for the same semantics can likely be improved further. Adding information on more visual variables has potential, but caution was encouraged lest the notation become too complex. There are other issues requiring attention that these revisions did not address. Some people will use all or some notational elements, but adoption would be higher with good software support.

8. Discussion

Study 1: Contextualization

Responses expand the information available on how graphical HTAs are used. Shepherd's (2001) comments on HTA as a communication tool within teams and with non-HFE staff are emphasised. The importance of fast mental-model building on the structure/hierarchy of tasks also appeared repeatedly. The results overall suggested that improvements were wanted, and provided a reference point to ensure revisions would improve real-world use, rather than just an arbitrary metric. A set of open questions was a highly successful format for response elicitation. Visualisation of plans and decisions were the issues most brought up

Use of Principles

This highlighted areas for improvement and formalised their interpretation.

Although the changes devised based on the principles appeared to be an improvement, repeated refinement with stakeholders was required to maximise cognitive effectiveness. It is uncertain whether successive applications of the principles would have led to the same result. This is perhaps due to the author's lack of experience applying them.

Study 2: Design Validation & Refinement

Without instruction, participants interpreted diagrams in very different ways, similar to Patrick *et al.* (2000).

The attempt to visually integrate tasks and subtasks was too successful for notations 1 and 2, so had to be diminished for notation 4. The results indicated that people with no prior experience with either notation preferred a visual representation of the plan. Notation 4 was the best of those considered, with the improvements in terms of reduced symbol deficit, high perceptual discrimination, additional scope for secondary notation, semantically transparent (for some) or translucent symbols; an open/closed metaphor for level of task description, consistent across tasks and encapsulation. It was thus worth testing the revisions against traditional HTA.

Study 3: Empirical Comparison

Accuracy

Significantly more questions were answered correctly with RN than TN, suggesting RN aids reader accuracy. This appeared to be due to a couple of specific elements rather than the combination of all of them: participants were significantly more accurate for questions on determining whether tasks necessarily took place. On the individual sub-questions, the difference between notations was significant only when the task did not occur, i.e. participants marked tasks as definite when they should have been potential. This suggests that users default to the position that if the task is on the HTA, it will necessarily occur as part of the task performance, and clearly not all of them understood the plan well enough to dissuade them from this for particular tasks. The dashed line appears to manage this where the plan does not, perhaps as a result of perceptual popout (Quinlan, 2003).

An argument could be made that the plan was not clear enough. This appears unlikely, considering a number of TN participants did answer these questions correctly, with nothing but the plan to guide them. Regardless, this is an inherent danger of a written plan: it can be worded in a manner difficult to understand for its intended audience.

For all participants or females alone, accuracy was insignificantly different on questions of hierarchy for RN. An improvement was significant when considering male participants. RN was at a disadvantage given that it had not been introduced/taught to the participants. That any improvements were made at all with the unfamiliar RN corroborates results from Peebles & Cheng (2003) that representational advantages can more than offset user familiarity.

Multiple improved measures had *p*-values between .05 and .1 (Appendix G), perhaps with more participants, significant differences would have been found. Also notable is that some questions were answered correctly by all or none of the participants in both conditions, negating the possibility of determining a difference. A different piloting strategy may be required.

Ease

Ease ratings for students (with one notable exception) were either insignificantly different or lower for the revised notation, regardless of higher scores. This contrasted with results from novices and experts. Four potential factors are suggested:

- TN was overrated: RN's diagrammatic complexity reflected the complexity of the task, making the difficulty more immediate. A lack of awareness of scope would explain some TN participants marking questions at the maximum level of ease despite getting them entirely or mostly wrong.
- RN was underrated: participants were unfamiliar with both HTA interpretation and RN. They had no information on RN's use other than the key. This was reflected in their comments. Novices were equally unfamiliar with all notations. HFE practitioners had sufficient experience with HTA to interpret the potential for use. This may have been different if participants had been taught RN initially.
- Higher visual complexity exaggerated the impression of difficulty.
- Even when all were the same between *n* tasks, *n-1* sequence symbols had to be interpreted (and checked against the key) in a linear fashion, particularly when determining task ordering. This is not dissimilar to reading a written plan.

Also notable is that participants had no 'objective' standard, and were not able to directly compare ease, as they could in a repeated-measures design experiment (Maxwell, 1978).

Speed

Faster accurate interpretation is an important improvement. There is a possible confounding variable of time taken in writing comments, but there do not appear to be more for TN. Time was only measured in minutes, which was coarse, but precise enough to find significant differences. It is uncertain which notational changes caused the speedup.

Gender difference

A gender difference in ability and preference was observed in the empirical study, with females lower than males for both measures, echoing cross-culturally consistent markers of higher visuospatial ability and preference in males (Halpern, 2013; Blackwell, 1997). Lower female preference did not appear in the expert evaluation. The difference may be accounted for by level of training, which can mitigate differences in spatial-visualization aptitude (Frandsen and Holder, 1969), or by few females in the latter study.

Study 4: Expert Evaluation

HFE practitioner responses provided another helpful viewpoint to compare with novices and students; these have been discussed in those sections. There was broad agreement on positives of the revised notation and need for software support, butmore discrepancy regarding use of additional visual variables and how they expected non-HFE task experts to respond to it.

The ‘subgroup ordering’ suggestion (Figure 20) would indicate the ordering for multiple tasks and group them with one syntax element. This should reduce visual noise and provide better cueing for which task sequences are complex (and so require more attention), while maintaining a level of expressiveness and clarity. Repeated notation (such as arrows) may be useful to avoid long line with no indication of sequencing type.

Use of colour was not thought helpful for coding a large number of types: the link between colour and code is (in the case of tools) semantically opaque, requiring repeated reference to the key. Colour becomes distracting or difficult to distinguish when multiple codes are needed for one task. One alternative suggested was to accompany tasks with semantically-related symbols; standardised versions could be used where they exist (BS ISO 7000:2014 for equipment symbols).

As a methodological note, most themes appeared in the first 5 participants, which mirrors suggestions for HCI/UX testing (Nielsen, 2000).

Although the author attempted to fairly represent the negative feedback on the notation, bias may have persisted. Furthermore, the themes were not independently assessed by a third party (Schreier, 2014).

Further research

Many comments were made by practitioners on the problems of complexity. Given that the two main strategies for complexity management are already central to HTA, one alternative presents itself: find shorthand semantics for representing common patterns in tasks that are approximately informationally-equivalent. One common suggestion from the contextualisation may be one of a number of steps toward this: depict task context, particularly regarding dependencies and information on how tasks and risk factors interact. More generically this can be considered how tasks affect, and are affected by, world state. Another level up in abstraction is task input/output. It is not clear *a priori* at what conceptual level the best approach here would be.

The hierarchical depiction introduced here provides enclosed space between task and subtasks for additional semantics, although care must be taken to not add complexity without commensurate return.

Additional research should:

– Test the use of RN for making and learning HTAs, to be comparable with Adams *et al.* (2013). This would provide better empirical evidence of changes in understanding of logical decomposition of hierarchy.
– Compare and integrate ‘revised’ and ‘simplified’/‘subgroup ordering’ notations.
– Evaluate the revised notation in a ‘real-world’ scenario, as used by practitioners.
– Determine differences between occupations in interpretation of diagrams and whether this requires any differentiation between representations.
– Investigate whether gender differences generalise to other populations.

9. Conclusions

Graphical HTAs are commonly used for communication in HFE teams and with non-ergonomists.

All three evaluative studies strongly indicated that graphical HTAs should notate non-definite tasks. The dashed line notation represents this well, providing both semantic transparency and high perceptual discriminability through unique visual texture.

The revised depiction of hierarchy - with its consistent ‘open/closed’ metaphor - aided semantic transparency of logical decomposition for novices, and was considered more clear by experts. It improved understanding of hierarchical structure for male students without negatively affecting females. It also increases the capacity for further improvements.

Gender differences were noted: males were more accurate than females and preferred the more visual notation, which females did not. Lower preference by females was not found for HFE practitioners.

The subjective feedback for plan sequencing notation reinforces that the notation is perceptually distinctive and semantically useful. The empirical data on this was positive but inconclusive.

The notation added some complexity, but this was mostly seen as a worthwhile trade-off by novices and experts. A suggestion to lower sequencing notation complexity with subgroup ordering is promising, and would benefit from further research. This may improve mediocre ease ratings, as would initial training.

Time per correct answer was lower with RN.

The project adds to the growing set of examples demonstrating the importance of graphical notation: subtle changes can have meaningful consequences.

10. References

Adams, A.E., Rogers, W.A. and Fisk, A.D., 2013. Skill components of task analysis. *Instructional Science*, *41*(6), pp.1009-1046.

Annett, J., Cunningham, D. and Mathias-Jones, P., 2000. A method for measuring team skills. *Ergonomics*, *43*(8), pp.1076-1094.

Annett, J., Duncan, K.D., 1967. Task analysis and training design. *Occup. Psych*. 12, 211–221.

Blackwell, A.F., 1997. Diagrams about thoughts about thoughts about diagrams. In *Reasoning with Diagrammatic Representations II: AAAI 1997 Fall Symposium* (pp. 77-84).

British Standards, 2014. BS ISO 7000: *Graphical symbols for use on equipment — Registered symbols.*

Brooke, J.B. and Duncan, K.D., 1981. Effects of system display format on performance in a fault location task. *Ergonomics*, *24*(3), pp.175-189.

Burton-Jones, A., Wand, Y. and Weber, R., 2009. Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems*, *10*(6), pp.495-532.

Carpendale, M.S.T. 2003. Considering Visual Variables as a Basis for Information Visualisation. *Computer Science TR #2001-693-16.* Dept. Computer Science, University of Calgary.

Chung, P.H., Zhang, J., Johnson, T.R. and Patel, V.L., 2003. An extended hierarchical task analysis for error prediction in medical devices. In *AMIA Annual Symposium Proceedings* (Vol. 2003, p. 165). American Medical Informatics Association.

Ciccarelli, F.D., Doerks, T., Von Mering, C., Creevey, C.J., Snel, B. and Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. *science*, *311*(5765), pp.1283-1287.

Cureton, E.E., 1956. Rank-biserial correlation. *Psychometrika*, *21*(3), pp.287-290.

Elo, S. and Kyngäs, H., 2008. The qualitative content analysis process. *Journal of advanced nursing*, *62*(1), pp.107-115.

Felipe, S.K., Adams, A.E., Rogers, W.A. and Fisk, A.D., 2010, September. Training novices on hierarchical task analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 23, pp. 2005-2009). SAGE Publications.

Figl, K., Mendling, J., and Strembeck, M., 2013. The influence of notational deficiencies on process model comprehension. *Journal of the Association for Information Systems* , *14* (6), p.312.

Frandsen, A.N. and Holder, J.R., 1969. Spatial visualization in solving complex verbal problems. *The Journal of Psychology*, *73*(2), pp.229-233.

Gagne, R.M., 1968. Presidential address of division 15 learning hierarchies. *Educational psychologist*, *6*(1), pp.1-9.

Galesic, M. and Bosnjak, M., 2009. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public opinion quarterly*, *73*(2), pp.349-360.

Gillan, D.J., Wickens, C.D., Hollands, J.G. and Carswell, C.M., 1998. Guidelines for presenting quantitative data in HFES publications. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(1), pp.28-41.

Green, T.R., 1989. Cognitive dimensions of notations. *People and computers V*, pp.443-460.

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

Gyi, D., Shalloe, S. and Wilson, J.R., 2015. Participatory ergonomics. *Evaluation of Human Work*, pp.883-906.

Halpern, D.F., 2013. *Sex differences in cognitive abilities*. Psychology press.

Hellier, E., Edworthy, J. and Lee, A., 2001. An analysis of human error in the analytical measurement task in chemistry. *International Journal of Cognitive Ergonomics*, *5*(4), pp.445-458.

Hitchman, S., 2002. The details of conceptual modelling notations are important-a comparison of relationship normative language. *Communications of the Association for Information Systems*, *9*(1), p.10.

Jenkins, D., Stanton, N., Walker, G., Salmon, P. and Young, M., 2006. Creating interoperability between the hierarchical task analysis and the cognitive work analysis tools. *Report from the Human Factors Integration Defence Technology Centre, UK*.

Kerby, D.S., 2014. The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation 1. *Comprehensive Psychology*, *3*, pp.11-IT.

Kirwan, B. and Ainsworth, L. 1992. *A Guide to task analysis*. London: Taylor & Francis.

Kurosu, M. and Kashimura, K., 1995, May. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference companion on Human factors in computing systems* (pp. 292-293). ACM.

Kuselman, I., Pennecchi, F., Fajgelj, A. and Karpov, Y., 2013. Human errors and reliability of test results in analytical chemistry. *Accreditation and Quality Assurance*, *18*(1), pp.3-9.

Kutar, M., Britton, C. and Jones, S., 1997. *A graphical representation for communicating sequential processes*. University of Hertfordshire.

Larkin, J.H. and Simon, H.A., 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, *11*(1), pp.65-100.

Leech, Nancy L., and Anthony J. Onwuegbuzie. "A Call for Greater Use of Nonparametric Statistics." In *Mid-South Educational Research Association Annual Meeting*, vol. 2002, no. 1. 2002.

LinkedIn, 2017a. *Chartered Institute of Ergonomics and Human Factors*. Available at: https://www.linkedin.com/groups/93258. [Accessed April 2017].

LinkedIn, 2017b. *Human Factors*. Available at: https://www.linkedin.com/groups/126397. [Accessed April 2017].

LinkedIn, 2017c. *Human Factors and Ergonomics Society (HFES)*. Available at: https://www.linkedin.com/groups/163326. [Accessed April 2017].

LinkedIn, 2017d. *Ergonomics*. Available at: https://www.linkedin.com/groups/126396. [Accessed April 2017].

Lohse, G.L., 1997. The role of working memory on graphical information processing. *Behaviour & Information Technology*, *16*(6), pp.297-308.

Maxwell, C., 1978. Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. *British journal of clinical pharmacology*, *6*(1), pp.15-24.

McDonald, J.H., 2009. *Handbook of biological statistics* (Vol. 2, pp. 173-181). Baltimore, MD: Sparky House Publishing.

Miller, C.A. and Vicente, K.J., 2001. Comparison of display requirements generated via hierarchical task and abstraction-decomposition space analysis techniques. *International Journal of Cognitive Ergonomics*, *5*(3), pp.335-355.

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), pp.756-779.

Moody, D. and van Hillegersberg, J., 2008, September. Evaluating the visual syntax of UML: An analysis of the cognitive effectiveness of the UML family of diagrams. In *International Conference on Software Language Engineering*(pp. 16-34). Springer Berlin Heidelberg.

Morgan, J. and Michaelson, G., 2012. A comparative evaluation of tabular and graphical presentation styles for information retrieval search results.

Nachar, N., 2008. The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, *4*(1), pp.13-20.

Nielsen, J. 2000. *Why You Only Need to Test with 5 Users*. Nielsen Norman Group. Available at: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [Accessed 11 Nov. 2015].

Ormerod, T.C. and Shepherd, A., 2003. Using task analysis for information requirements specification: the Sub-Goal Template (SGT) method. *The handbook of task analysis for human-computer interaction*, p.347.

Patrick, J., Gregov, A. and Halliday, P., 2000. Analysing and training task analysis. *Instructional science*, *28*(1), pp.51-79.

Peebles, D. and Cheng, P.C.H., 2003. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(1), pp.28-46.

Petre, M. (1995). Why looking isn't always seeing: readership skills and graphical programming. *Communications of the ACM*, *38*(6), 33-44.

Quinlan, P.T., 2003. Visual feature integration theory: past, present, and future. *Psychological bulletin*, *129*(5), p.643.

Rauterberg, M., 1999. Perception, Cognition, Action: an Action Theoretical Approach. *Perception*, *14*(1), pp.1-8.

Alexander, J., Little, K. and Reece, A., 2017. *User-centred Leatherman Redesign Logbook*. Undergraduate. Loughborough University.

Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D., Ladva, D., Rafferty, L. and Young, M., 2009. Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, *39*(3), pp.490-500.

Schreier, M., 2014. Qualitative content analysis. *The SAGE handbook of qualitative data analysis*, pp.170-183.

Sebillotte, S., 1988. Hierarchical planning as method for task analysis: the example of office task analysis. *Behaviour & Information Technology*, *7*(3), pp.275-293.

Shepherd, A., 1985. Hierarchical task analysis and training decisions. *Programmed Learning and Educational Technology*, *22*(2), pp.162-176.

Shepherd, A. 2001. *Hierarchical task analysis*. London: Taylor & Francis.

Sawilowsky, S.S., 2005. Misconceptions Leading to Choosing the t Test Over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter. *Journal of Modern Applied Statistical Methods*, *4*(2), p.26.

Stanton, N.A., 2006. Hierarchical task analysis: Developments, applications, and extensions. *Applied ergonomics*, *37*(1), pp.55-79.

Stanton, N.A. and Stevenage, S.V., 1998. Learning to predict human error: issues of acceptability, reliability and validity. *Ergonomics*, *41*(11), pp.1737-1756.

Stanton, N.A., Salmon, P., Harris, D., Marshall, A., Demagalski, J., Young, M.S., Waldmann, T. and Dekker, S., 2009. Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach. *Applied ergonomics*, *40*(3), pp.464-471.

Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 115-122). ACM.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. Decision Sciences, 22(2), 219-240.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. Information systems research, 2(1), 63-84.

Waddingham, A., 2014. *New Hart's Rules: The Oxford Style Guide*. Oxford University Press (UK).

Wendt, H.W., 1972. Dealing with a common problem in Social science: A simplified rank-biserial coefficient of correlation based on the U statistic. *European Journal of Social Psychology*, *2*(4), pp.463-465.

Wickens, C.D. and Carswell, C.M., 1995. The proximity compatibility principle: its psychological foundation and relevance to display design. *Human factors*, *37*(3), pp.473-494.

Wilson, J.R. and Rajan, J.A., 1995. Human-machine interfaces for systems control. *Evaluation of human work: a practical ergonomics methodology*, pp.357-405.

Appendix A: Participation Information

Investigator

Andrew Reece - a.z.m.reece-14@student.lboro.ac.uk

Supervisor

Simon Hodder - S.Hodder@lboro.ac.uk

Environmental Ergonomics Research Centre, Loughborough Design School, Loughborough University, LE11 3TU

01509 223685

What is the purpose of the study?

Hierarchical task analysis is a way of breaking down the steps involved in a task, that was introduced in the late 1960s. Despite progress in related fields, there has been little substantial change to the visual representation of these analyses. This study is one of a number of steps to determine which features should be changed and evaluate the difference between the original and modified versions of the visual representations of the analysis.

Who is doing this research and why?

This study is part of a student research project supported by Loughborough University.

Are there any exclusion criteria?

No.

What will I be asked to do?

Answer questions relating to your use of hierarchical task analysis

Answer questions based on a hierarchical task analysis given to you. The questions will be on the task analysed, your opinions on the representation, and brief comparison with another representation.

Answer questions relating to your thoughts on a graphical representation of hierarchical task analysis

Once I take part, can I change my mind?

Yes. After you have read this information and asked any questions you may have we will ask you to complete an Informed Consent Form, however if at any time, before, during or after the sessions you wish to withdraw from the study please just contact the main investigator. You can withdraw at any time, for any reason and you will not be asked to explain your reasons for withdrawing.

However, once the results of the study are aggregated/published/submitted has been submitted (expected to be by June 2017), it will not be possible to withdraw your individual data from the research.

Will I be required to attend any sessions and where will these be?

No, you can answer by phone or email.

How long will it take?

It depends on how much detail you go into. It should not take longer than 30 minutes, unless you wish to elaborate.

What personal information will be required from me?

The job context in which you use HTA.

Are there any risks in participating?

None foreseen.

Will my taking part in this study be kept confidential?

Yes, your data will be anonymised, and any identifying information will be deleted once the study is finished (expected June 2017).

I have some more questions; who should I contact?

Contact either the investigator or the supervisor. The contact details for both are at the top of this document.

What will happen to the results of the study?

The results will be used in a student dissertation. There is also the possibility they will be used in a conference or journal publication.

What if I am not happy with how the research was conducted?

If you are not happy with how the research was conducted, please contact Ms Jackie Green, the Secretary for the University's Ethics Approvals (Human Participants) Sub-Committee:

Ms J Green, Research Office, Hazlerigg Building, Loughborough University, Epinal Way, Loughborough, LE11 3TU. Tel: 01509 222423. Email: J.A.Green@lboro.ac.uk

The University also has a policy relating to Research Misconduct and Whistle Blowing which is available online at http://www.lboro.ac.uk/committees/ethics-approvals-human-participants/additionalinformation/codesofpractice/ .

Appendix B: Consent Form

Modernizing the graphical representation of hierarchical task analysis
INFORMED CONSENT FORM
(to be completed after Participant Information Sheet has been read)

Taking Part	**Please initial box**
The purpose and details of this study have been explained to me. I understand that this study is designed to further scientific knowledge and that all procedures have been approved by the Loughborough University Ethics Approvals (Human Participants) Sub-Committee.	☐
I have read and understood the information sheet and this consent form.	☐
I have had an opportunity to ask questions about my participation.	☐
I understand that I am under no obligation to take part in the study, have the right to withdraw from this study at any stage for any reason, and will not be required to explain my reasons for withdrawing.	☐
I agree to take part in this study. Taking part in the project may include being photographed, interviewed and recorded (audio or video).	☐
Use of Information	
I understand that all the personal information I provide will be treated in strict confidence and will be kept anonymous and confidential to the researchers unless (under the statutory obligations of the agencies which the researchers are working with), it is judged that confidentiality will have to be breached for the safety of the participant or others or for audit by regulatory authorities.	☐
I understand that anonymised quotes may be used in publications, reports, web pages, and other research outputs.	☐
I agree for the data I provide to be securely archived at the end of the project.	☐

Name of participant	[printed]	Signature	Date
Researcher	[printed]	Signature	Date

Appendix C: Contextualization questions

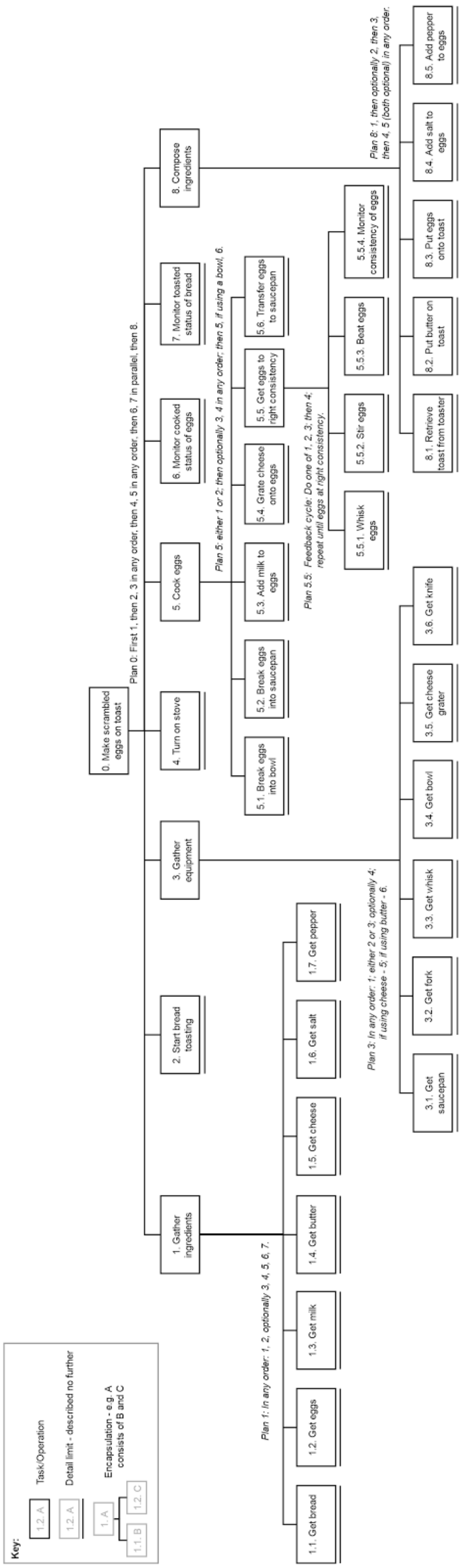
1) Why do you use HTAs?
2) For what tasks do you use HTAs? At what level of abstraction/detail are they?
3) What is your process for creating HTAs?
4) For what purpose do you use the graphical representation of HTAs?
5) For what purpose do you use the tabular representation of HTAs?
6) Which representation do you spend more time considering?
7) What do you find difficult/awkward about the graphical format?
8) What do you find difficult/awkward about the tabular format?
9) What type of information would you find helpful on a graphical representation?

10a) What other task representation techniques (if any) do you use in conjunction with HTA?

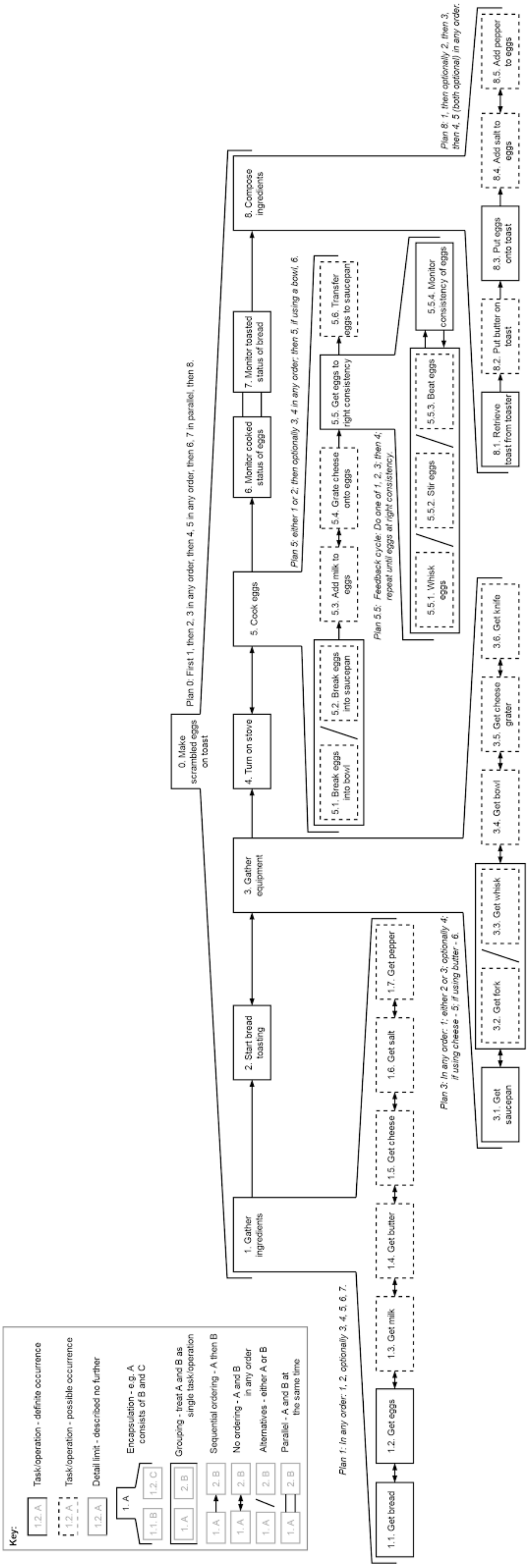
10b) How does HTA complement these other techniques (or not)?

11) How large/complex are your HTAs? If large/complex, how do you deal with this?
12) How complex do your 'plans' tend to be? Do they require flowcharts?

Appendix D: Scrambled egg task, TN



Appendix E: Scrambled egg task, RN



Appendix F: Question Sheet

Please note the time when you start: ________________

Section A - Task-specific questions

1. According to the HTA, will the following operations definitely occur? (Y/N)
   a. Get eggs ___________
   b. Get bowl ___________
   c. Stir eggs ___________
   d. Retrieve toast from toaster ___________
   e. Transfer eggs to saucepan ___________

Complete the following statement by making a mark on the subsequent scale:

The diagram made determination of this information...

Very difficult *Very easy*

|________________________________|________________________________|

Additional comments:

2. For each of the following, name all of the possible orders of A and B (assuming both are done), e.g. '*A then B...' :*
   a. A) Gather ingredients, B) Start bread toasting

      ____

   b. A) Get bread, B) Get eggs

      ____

   c. A) Turn on stove, B) Add milk to eggs

      ____

   d. A) Prepare eggs, B) Get saucepan

      ____

The diagram made determination of this information...

Very difficult *Very easy*

|________________________________|________________________________|

Additional comments:

3. Which operations, if any, occur at the same time? ____________________

The diagram made determination of this information...

Very difficult *Very easy*

|_________________________________|_________________________________|

Additional comments:

4. Please circle the hierarchical relationship between the following operations (where ‘Make scrambled eggs on toast’ is the highest point in the hierarchy):
    a. ‘Get milk’ is **higher/the same/lower** in the hierarchy compared to ‘Put butter on toast’
    b. ‘Break eggs into saucepan’ is **higher/the same/lower** in the hierarchy compared to ‘Monitor cooked status of eggs’
    c. ‘Stir eggs’ is **higher/the same/lower** in the hierarchy compared to ‘Put butter on toast’

The diagram made determination of this information...

Very difficult *Very easy*

|_________________________________|_________________________________|

Additional comments:

5. How many (lowest level) operations are strictly necessary to make scrambled eggs on toast? (as described here) ____________

The diagram made determination of this information...

Very difficult *Very easy*

|___________________________________|___________________________________|

Additional comments:

6. Mention one of the operations that is at the greatest level of detail in this diagram: ____________

The diagram made determination of this information...

Very difficult *Very easy*

|___________________________________|___________________________________|

Additional comments:

7. Overall, the task analysis was … to understand:

Very difficult *Very easy*

|___________________________________|___________________________________|

Additional comments:

Please note the time when you finish Section A: __________________

Section B - Questions on graphical representation

1. **Semantic transparency - understandability of symbols**. *Symbols are:*
    a. ***Transparent/immediate*** *if novice readers would be able to infer its meaning just from the way it looks (e.g. stick figure → person).*
    b. ***Opaque*** *if the relationship between appearance and meaning is arbitrary. The symbol has no suggestion of its meaning.*
    c. ***Perverse*** *if the appearance suggests a different/opposite meaning than the one intended.*
    d. ***Translucent*** *(between transparency and opacity) if symbols suggest their meaning (helpful as a mnemonic) but require initial explanation.*

For each symbol (e.g. as shown on the key), please mark the corresponding level of semantic transparency:

Task:

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Encapsulation/Hierarchy ('... consists of …' / 'described in more detail as …'):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Grouping ('consider ... and … together'):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Sequential ordering ('... then …'):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Non-sequential ordering ('... and … in any order'):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Selection ('either … or …'):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Parallel ('... and … at the same time' / concurrently):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Definite occurrence:

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Uncertain occurrence:

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

Detail limit (described no further):

Perverse *Opaque* *Transparent*

|________________________________|________________________________|

2. **Perceptual discriminability of symbols.** Were there any sets of 2 or more symbols that were not clearly distinguishable from each other?

If yes, please elaborate below:

3. **Diagram complexity** - how complex did you find the diagram of this task?

Not at all complex *Very complex*

|____|

Additional comments:

4. **Notation complexity** - how complex did you find the set of symbols used to represent tasks more generally?

Not at all complex *Very complex*

|____|

Additional comments:

5. **Representation suitability** - how suitable do you find this representation of tasks?

Not at all suitable *Very suitable*

|____|

Are there any groups of people that might use HTAs who you think would find it significantly more or less suitable than your answer? Please mention any below:

Additional comments:

Section C - Comparative questions

After everyone has finished you will be shown an alternative representation of the same task. Taking into account your answers in Section B, how would you rate this second representation in the following axes?

1. **Diagram complexity** - how complex did you find the diagram of this task?

Not at all complex *Very complex*

|____|

Additional comments:

2. **Notation complexity** - how complex did you find the set of symbols used to represent tasks more generally?

Not at all complex *Very complex*

|____|

Additional comments:

3. **Representation suitability** - how suitable do you find this representation of tasks?

Not at all suitable *Very suitable*

|____|

Additional comments:

4. **Preferred representation to read** - which of the 2 representations you have seen today do you prefer to read?

Initial representation *No preference* *Second representation*

|________________________________|________________________________|

What are the reasons behind this preference?

5. **Preferred representation to make** - which of the 2 representations you have seen today would you prefer to make?

Initial representation *No preference* *Second representation*

|________________________________|________________________________|

What are the reasons behind this preference?

If you have any additional thoughts/comments, or you can think of improvements to the task representations, please include them below. Feel free to sketch or write as appropriate, and ask for more paper if needed:

Appendix G: Mann-Whitney U Results

Question(s)	n [TN]	n [RN]	U	rank-biserial	p (1-tailed)	p (2-tailed)
All Participants						
Section A	15	14	56.5	0.462	0.016	
Occurrence (1)	15	14	47	0.552	0.003	
Ordering (2)	15	14	97.5	0.071	0.377	
Parallel (3)	15	14	80	0.238	0.18	
Hierarchy (4)	15	14	79	0.248	0.132	
Abs Delta (5)	15	14	74.5	0.290	0.31	
Sequential	15	14	81	0.229	0.146	
Non-sequential	15	14	75	0.286	0.099	
1a	15	14	105	0.000	1	
1b	15	14	50	0.524	0.006	
1c	15	14	57	0.457	0.014	
1d	15	14	97.5	0.071	0.483	
1e	15	14	63.5	0.395	0.022	
2a	15	14	97	0.076	0.473	
2b	15	14	78.5	0.252	0.15	
2c	15	14	88.5	0.157	0.298	
2d	15	14	81	0.229	0.177	
4a	15	14	99.5	0.052	0.542	
4b	15	14	79.5	0.243	0.175	
4c	15	14	88	0.162	0.3	
6	15	14	89	0.152	0.291	
Time/Correct Ans.	14	13	52.5	0.423	0.031	
Ease Q2	13	12	28.5	0.635		0.006
Ease Q4	15	14	69.5	0.338		0.125
Ease Q6	15	14	79	0.248		0.265

Male	N [TN]	N [RN]	U	rank-biserial	p (1-tailed)	p (2-tailed)
Section A	9	7	12	0.619	0.018	
Occurrence (1)	9	7	20	0.365	0.121	
Ordering (2)	9	7	26.5	0.159	0.391	
Parallel (3)	9	7	16	0.492	0.072	
Hierarchy (4)	9	7	13	0.587	0.033	
Sequential	9	7	30.5	0.032	0.7	
Non-sequential	9	7	25.5	0.190	0.301	
1a	9	7	31.5	0.000	1	
1b	9	7	22	0.302	0.202	

1c	9	7	22	0.302	0.202	
1d	9	7	31.5	0.000	1	
1e	9	7	21	0.333	0.15	
2a	9	7	31.5	0.000	1	
2b	9	7	30	0.048	0.635	
2c	9	7	24	0.238	0.329	
2d	9	7	30.5	0.032	0.7	
A4a	9	7	28.5	0.095	0.549	
A4b	9	7	24	0.238	0.329	
A4c	9	7	15	0.524	0.055	
A6	9	7	26	0.175	0.338	
Ease Q3	9	7	12.5	0.603		0.045
Female	N [TN]	N [RN]	U	rank-biserial	p (1-tailed)	p (2-tailed)
Section A	6	7	9	0.571	0.048	
Occurrence (1)	6	7	6	0.714	0.015	
Ordering (2)	6	7	19	0.095	0.459	
Paralle (3)	6	7	19.5	0.071	0.617	
Hierarchy (4)	6	7	19	0.095	0.413	
Sequential	6	7	13.5	0.357	0.182	
NonSequential	6	7	9	0.571	0.066	
1a	6	7	21	0.000	1	
1b	6	7	3	0.857	0.004	
1c	6	7	6.5	0.690	0.025	
1d	6	7	18	0.143	0.538	
1e	6	7	10	0.524	0.086	
2a	6	7	18.5	0.119	0.563	
2b	6	7	10	0.524	0.086	
2c	6	7	18	0.143	0.538	
2d	6	7	13	0.381	0.209	
4a	6	7	17.5	0.167	0.462	
4b	6	7	16.5	0.214	0.413	
4c	6	7	17.5	0.167	0.462	
6	6	7	18.5	0.119	0.563	

Appendix H: Additional RN HTA