Published in final edited form as:
Instr Sci. 2013 November ; 41(6): 1009–1046. doi:10.1007/s11251-013-9270-9.

Skill components of task analysis

Anne E. Adams, **Wendy A. Rogers**, and **Arthur D. Fisk**
School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170, USA

Abstract

Some task analysis methods break down a task into a hierarchy of subgoals. Although an important tool of many fields of study, learning to create such a hierarchy (redescription) is not trivial. To further the understanding of what makes task analysis a skill, the present research examined novices' problems with learning Hierarchical Task Analysis and captured practitioners' performance. All participants received a task description and analyzed three cooking and three communication tasks by drawing on their knowledge of those tasks. Thirty six younger adults (18–28 years) in Study 1 analyzed one task before training and five afterwards. Training consisted of a general handout that all participants received and an additional handout that differed between three conditions: a list of steps, a flow-diagram, and concept map. In Study 2, eight experienced task analysts received the same task descriptions as in Study 1 and demonstrated their understanding of task analysis while thinking aloud. Novices' initial task analysis scored low on all coding criteria. Performance improved on some criteria but was well below 100 % on others. Practitioners' task analyses were 2–3 levels deep but also scored low on some criteria. A task analyst's purpose of analysis may be the reason for higher specificity of analysis. This research furthers the understanding of Hierarchical Task Analysis and provides insights into the varying nature of task analyses as a function of experience. The derived skill components can inform training objectives.

Keywords

Task analysis; Training; Skill acquisition; Skill components

Introduction

Some task analysis (TA) methods are used to understand, discover, and represent a task in terms of goals and subgoals, for example, Hierarchical Task Analysis (HTA, Annett and Duncan 1967) and Goal-Directed Task Analysis (GDTA, Endsley et al. 2003). Although widely described in procedure and underlying skills (e.g., Crandall et al. 2006; Kirwan and Ainsworth 1992), there is still much to be learned. The few existing training studies of HTA we found indicated that learning HTA, for example, is not trivial (e.g., Patrick et al. 2000; Stanton and Young 1999). The present research approached TA as a skill acquisition problem to be understood through scientific inquiry. Two studies were designed to characterize novice and experienced TA performance and identify skill components.

Correspondence to: Anne E. Adams.

Electronic supplementary material The online version of this article (doi:10.1007/s11251-013-9270-9) contains supplementary material, which is available to authorized users.

Task analysis (TA)

TA refers to a collection of methods used in wide variety of areas. In general, a task analyst aims to understand a task/work, its context, and performance for the purpose of improving or supporting effectiveness, efficiency, and safety via design or training (Annett 2004; Diaper 2004; Hoffman and Militello 2009; Jonassen et al. 1989; Kirwan and Ainsworth 1992; Redish and Wixon 2003). To illustrate, TA is used in the nuclear and defense industry (Ainsworth and Marshall 1998), has helped evaluate food menu systems (Crawford et al. 2001), identify errors in administering medication (Lane et al. 2006), assess shopping and phone skills in children with intellectual disabilities (Drysdale et al. 2008), and furthered understanding of troubleshooting (Schaafstal et al. 2000).

The general process of TA is iterative and involves planning and preparing for the TA, gathering data, organizing, and analyzing them. It involves reporting (i.e., presenting findings and recommending solutions) and verifying the outcome (e.g., Ainsworth 2001; Clark et al. 2008; Crandall et al. 2006). TA methods differ depending on which phase(s) of TA they focus on, what aspects of a task they target (e.g., behaviors, knowledge, reasoning), how results are presented, and what type of recommendations result.

HTA and hierarchical organization

HTA, used over 40 years, is admittedly difficult to learn (e.g., Diaper 2004; Stanton 2006). Data organization in HTA occurs through two processes: *redescription* and *decomposition*. During redescription the analyst defines the *main goal* (task description) and breaks it down ("re"-describes it) into lower-level *subgoals*, recursively, until reaching a predetermined stopping criterion. This results in a *hierarchy* (HTA diagram) as shown in Fig. 1 that provides a task overview and serves as input/framework for further analysis. During *decomposition*, the analyst inspects each lower level subgoal with respect to categories such as input required, resulting action, feedback received, time to completion, or errors observed (Ainsworth 2001; Shepherd 1998), the result of which is better represented in a tabular format (Shepherd 1976).

Redescription is not trivial (Stanton 2006). Sub-Goal Template (SGT, Ormerod and Shepherd 2004; Shepherd 1993) provides templates to solve this problem but refers the analyst back to redescription when no template is available. Few studies specifically investigated novices' problems when learning how to apply HTA. They found, for example, that novices who just received training, tended to redescribe a task in terms of specific actions rather than subgoals (Patrick et al. 2000; Stanton and Young 1999), a similar challenge that novices learning GDTA encounter (Endsley et al. 2003). Difficult is also redescribing a higher-level subgoal equivalently into lower-level subgoals (e.g., not just *one* subgoal), knowing what subgoals to include/exclude, and specifying plans at *all* levels of analysis (Patrick et al. 2000). *Plans* (Fig. 1) specify the sequence and conditions of accomplishing subgoals (Ainsworth 2001).

Addressing the training need

Studies of training HTA pointed out general training needs, but the question remains specifically how to train redescription and what performance looks like across a number of tasks and levels of proficiency. TA is a complex cognitive skill (e.g., Patrick et al. 2000), and the 4C/ID-Model of instructional design suggests creating a hierarchy of constituent skills as the first of four activities in a principled skill decomposition (van Merriënboer 1997). Therefore, our overarching question was what component skills underlie TA, and particularly, hierarchical redescription?

Understanding skill components will help determine training objectives, assess performance, and tailor learner feedback. This is important as the need for training TA and specific methods is likely to increase, given the call for more (and more competent) practitioners and new TA methods (Crandall et al. 2006). Further training is also essential because a hierarchical redescription of subgoals feeds into subsequent TA methods (e.g., SHERPA, TAFEI, GOMS). A better understanding of the skill will advance the discussion on how to assess the quality of a TA.

To accomplish our research goal, we chose two levels of TA experience, six tasks for repeat measurement, one TA method to train novices (HTA), focusing on understanding redescription. Data from novices identify barriers to initial progress. Data from practitioners using TA in their job inform goals for skill development. Comparing the two delineate skill components because of their absence in novice performance (Seamster et al. 2000).

Tasks to be analyzed

The wide usage of TA makes it challenging to select tasks that are pertinent to all areas. Previous HTA training tasks included painting a door, making a cup of tea (Patrick et al. 2000), evaluating a radio-cassette machine (Stanton and Young 1999), as well as making a piece of toast, a cup of coffee, a phone call, and a South African main dish (Felipe et al. 2010). We reviewed tasks used in literature (Craik and Bialystok 2006; Davis and Rebelsky 2007; Felipe et al. 2010; Patrick et al. 2000; Shepherd 2001), and developed a set of six tasks (see Table 1). We chose two familiar domains (cooking, communication) to allow novices to focus on redescription. Not having to learn about a new domain and extract knowledge from subject-matter expert at the same time should reduce novices' intrinsic cognitive load (high degree of complexity; cf. Carlson et al. 2003). Moreover, being familiar with a task procedure could influence the availability of task-related information (Patrick et al. 2000); thus, each domain had one task for which the procedure was specific, general, or unknown (unfamiliar tasks).

Overview of research

To determine skill components in redescription we conducted two studies. Novices participated in Study 1 and practitioners in Study 2. Years of experience using TA served as a proxy for practitioner's proficiency. The goals of Study 1 were to characterize novices' HTA (product and process) and determine the effectiveness of three types of training. The goals of Study 2 were to characterize practitioners' TA (product and process). All participants analyzed six tasks. Novices analyzed one task before and five after training, whereas practitioners analyzed tasks while thinking aloud. Questionnaires assessed declarative knowledge (Study 1) and strategic knowledge (Study 1 + 2).

Study 1: novices

As background to characterizing novices' performance and determining the effectiveness of three types of training, we considered the scenario in which a novice reads an overview of HTA and applies that knowledge shortly thereafter. Although there is the occasional course devoted to task analysis (Crandall et al. 2006) and some novices receive initial training followed by months of expert mentoring (Stanton 2006; Sullivan et al. 2007), current methods of learning often include using books, web resources, or brief workshops (Crandall et al. 2006). Therefore, it was worthwhile to assess the limits of short declarative training procedures.

Declarative training

This study builds on Felipe et al. (2010) who employed two sets of instructions: one for all participants and one that differed across experimental conditions. Previous literature on training HTA used custom-made instructions (Patrick et al. 2000; Stanton and Young 1999). This study used declarative training in the form of an introduction to HTA, outlining the main concepts, that was available from literature (Shepherd 2001). Given novices' problems adhering to the HTA format (Patrick et al. 2000), we removed references to HTA format (hierarchy and tabular format) to learn what formats novices naturally choose to represent their analysis rather than assessing how well novices adhered to the specific HTA format.

Instructions on HTA usually include a flowchart or list of steps (e.g., Shepherd 2001). We compared the relative benefits of three types of additional instructions that emphasize different aspect of conducting HTA (procedures, decisions, and goals), illustrated by three types of spatial diagrams: matrix, network, and hierarchy. Figure 2 shows a list of steps exemplifying a *matrix* that statically relates element pairs in rows and columns. The decision-action-diagram is a *network* example, depicting dynamic information as a graph or path diagram. A concept map represents a *hierarchy*, that is, a tree diagram with information rigidly organized in nodes and links at different levels (Novick 2006; Novick and Hurley 2001).

We expected participants' HTA to reflect the instructions' emphasis on procedures, decisions, and goals. Novices who receive step-by-step (procedural) instructions on how to conduct HTA do not need to generate their own. Given novices' focus on lower levels of analysis (Patrick et al. 2000), these HTA were expected to do exactly that, specifying one task procedure. Participants receiving information about goals of HTA (concept map) have to generate their own procedures on how to conduct HTA, and thus we expected their HTA to contain more (higher-level) subgoals (as found by Felipe et al. 2010) and be more general. The decision-action diagram of HTA then should fall in between these two conditions.

Research on procedural training (how to do something) versus conceptual training (drawing attention to concepts and why) suggests an immediate benefit of procedural training on trained tasks but a benefit of conceptual training for novel tasks (e.g., Dattel et al. 2009; Hickman et al. 2007; Olfman and Mandviwalla 1994). In this study, analyzing an unfamiliar task constitutes a novel task. Given the absence of specific task information, we expected that all participants would produce general HTA; possibly more so participants receiving the concept map.

Evaluating novices' HTA

Novices' HTA were characterized on seven categories: format; breadth and depth; subgoals; plan; main goal; criteria; and versatility. First, we expected participants to use lists and flowcharts in absence of HTA format, given novices' tendency to use list and flowcharts even when instructed to use a hierarchy (Patrick et al. 2000). Second, we quantified the dimensions of HTAs through *depth* (the number of levels) and *breadth* (the number of subgoals making up the highest level), because hierarchical means that an HTA is at least two levels deep and literature provides guidance regarding HTA breadth. Rules of thumb regarding breadth include four and five subgoals (Stanton and Young 1999), no more than seven (Ainsworth 2001), between three and ten (Stanton 2006), or between four and eight (Patrick et al. 1986, as cited by Stanton 2006). We chose a breadth between three and eight subgoals to be most consistent with all recommendations.

The third question was whether participants recognized the importance of subgoals to HTA and what subgoals were identified. A subgoal was a verb-noun pair, for example, "obtain bread" and compared to a master HTA created for each task. Fourth, we were interested in whether participants would recognize the importance of a plan. Fifth, we assessed whether the participant stated the main goal (e.g., making a phone call), which provides important context for the HTA. Sixth was whether participants included satisfaction criteria that define conditions for determining whether a task was completed satisfactorily. Last, we determined if the HTA was versatile (general) and accounted for at least three task variations, for example, using different types of cell phones or a rotary phone.

Overview of Study 1

The goals of this study were to characterize novices' HTA (products and process), and determine the effectiveness of three types of training. Novices analyzed one task before and five tasks after training, which allowed assessment of naïve understanding of TA as well as trained performance. Training consisted of an introduction to HTA and a handout that different between training conditions. Procedural knowledge (HTA products) was assessed on seven characteristics. Declarative knowledge was assessed via a recall test, and strategic knowledge elicited via a questionnaire.

Method for novices

Participants—We report data collected from 11 male and 25 female undergraduate students. Participants ranged in age from 18 to 24 years ($M = 20.6$ years, SD = 1.5) with the majority being Caucasian (75 %). Participants' majors reflected the variety of majors offered at a large research university. Table 2 shows descriptive data and that participants did not significantly differ in their general abilities (measures of perceptual speed, working memory, and vocabulary). The experiment lasted approximately 2 h for which participants received two extra course credits.

Participants had to fulfill two criteria to be considered a novice. First, their initial TA was not rendered in the hierarchical or tabular format as prescribed by HTA. Second, participants had to report having no experience conducting a TA outside of class, as assessed by three questions in the *Demographics and Experience Questionnaire*. Although 25 % of the participants had heard about TA in a class, their initial TA was not in HTA format so they were included.

Tasks to be analyzed—Table 1 shows the six tasks to be analyzed and their range of expected degree of familiarity (low, high) and procedural specificity (specific, general). Tasks were simple enough for a draft to be completed within a short period (15 min).

Training materials—All participants received a three-page handout providing a general *Introduction to Hierarchical Task Analysis* adapted from Shepherd (2001). It included a brief overview of the history and goals of HTA and main concepts such as hierarchical nature, goals, subgoals, constraints, and plans for accomplishing the goal.

Participants in each training condition received an additional one-page handout with *Condition-Specific Instructions* shown in Fig. 2. These additional instructions focused on redescribing a higher-level goal into lower-level subgoals, but also included initial TA activities such as defining the purpose of the analysis and gathering data. In the Steps Condition, the additional information was presented as a bulleted list and focused on the sequence of steps (adapted from Stanton 2006, p. 62ff). The Decision-Action Diagram Condition provided a diagram illustrating the flow of decisions and actions (taken from Shepherd 1985). Concept Map Condition contained the information rendered as a concept map, including high-level goals of HTA (based on Shepherd 2001). To ensure that all participants were exposed to the same topics, information about determining if the redescription was equivalent was added to the Steps Condition, and the Decision-Action Diagram was amended with information about defining the purpose of the analysis and gathering data.

Questionnaires—Participants completed three questionnaires. The *Demographics and Experience Questionnaire* collected data on age, gender, education, and TA experience. The *Task Questionnaire* assessed familiarity with each task analyzed in the study (1 = not very familiar, 5 = very familiar) and how often those tasks were performed in everyday life (1 = never, 5 = daily). The *Task Analysis Questionnaire* assessed declarative and strategic knowledge that participants gained about HTA. Declarative knowledge was assessed by prompting participants to list and briefly describe the main features of HTA. Strategic knowledge was elicited by seven open-ended questions about how participants had identified goals and subgoals, indicated order, decided on breath and depth of the analysis, and what elements to analyze further. The *Task Analysis Questionnaire* also asked participants to rate difficulty and confidence of each task analysis; however, those data are not presented here.

Procedure—Participants read the informed consent and completed the ability tests listed in Table 2. To obtain a baseline measure for comparison after training participants received a written task description (as listed in Table 1) and were asked to perform a TA of either *making sandwich* or *making phone call*. Participants were free to use 11 × 17 in. paper as needed. Participants were not given a specific purpose for the TA. If participants had question, they were directed to work to the best of their knowledge and understanding of what it means to perform a TA. The experimenter collected the paper (HTA product) when participants put down their pen/pencil to indicate that they were done or when 15 min had passed. After the initial TA, participants received the *Introduction to Hierarchical Task Analysis* and had 10–15 min time to familiarize themselves with it. Then participants received the *Condition-Specific Instructions* for their training condition and were required to spend at least 5 min with this extra material but had up to 15 min available. Participants were allowed to make notes on the instructions.

After the instruction phase, participants analyzed two more tasks of the same domain following the same procedure as described above. Participants then completed the *Demographics and Experience Questionnaire* and contact information sheet before analyzing the three tasks of the second domain. Participants had 15 min for each analysis and could refer to the instructions throughout. The experimenter collected all instructions after the last HTA, and participants completed the *Task Questionnaire* and *Task Analysis Questionnaire* before being debriefed.

Design—This experiment was a between participant design with three training conditions: Steps, Decision-Action Diagram, and Concept Map. Task was a repeated measure (participants analyzed six tasks). Domain order (cooking, communication) was counterbalanced. Within a domain, task order was fixed as listed in Table 1; half the participants analyzed *making sandwich* (*making a phone call*) as Task 1 (before training) and the other half as Task 4 (after training). Participants were individually tested and randomly assigned to one training condition and counterbalance version. Procedural knowledge was assessed by coding participants' HTA. Declarative knowledge was determined via the first question of the *Task Analysis Questionnaire* (list five main features of HTA). The HTA process (strategic knowledge) was assessed by answers about decisions factor questions in the *Task Analysis Questionnaire*.

Results for novices

Data analysis addressed three questions: What are the HTA product characteristics before and after training on the seven criteria? Was there a beneficial effect of training? What strategies characterize the HTA process?

Task familiarity—A repeated measure ANOVA (task by condition by version) confirmed that familiarity ratings differed between tasks (main effect of task, $F = 711.79$, $df = 1.7$, $p < .01$, $\eta_p^2 = .96$) but not training conditions ($p = .14$) or counterbalance version ($p = .89$). The reported F value and degrees of freedom are Greenhouse-Geisser corrected. As expected, participants were very familiar with *making sandwich* and *making phone call* (high familiarity, Median = 5, range = 0) and unfamiliar with *making Vetkoek* and *sharing pictures using Adgers* (low familiarity, Median = 1, range = 0). Intermediate to high familiarity ratings emerged for *making breakfast* (Median = 5, range = 1) and *arranging meeting* (Median = 4, range = 4). Frequency ratings were in line with familiarity ratings: high for the high-familiarity tasks and low (never) for low-familiarity tasks. Thus, the task manipulation was successful.

Coding scheme—Table 3 shows the coding scheme for assessing novices' procedural and declarative knowledge. The categories were derived from Patrick et al. (2000) and the *Introduction to Hierarchical Task Analysis* that participants received. Two coders coded all material to ensure consistency of coding. Disagreements were resolved through discussion. Overall coder agreement for TA products was 79 % (range: 74–85 %; mean Cohen's Kappa = .73, range .68–.81). The total number of HTA features listed in the *Task Analysis Questionnaire* and included in data analysis was 169 (180 total—5 blanks—6 duplicates). Overall coder agreement for declarative knowledge was 85 % (Kappa = .81).

Format of HTA—When examining what format participants would naturally choose to render their TA, we expected to find lists and flowcharts. Data showed that the most common format was a list (58.3 %), usually a numbered list. Participants also used a flowchart (8.3 %), pictures such as shown in the right panel of Fig. 3 (11.1 %), or a combination of the above (22.2 %). One participant even initially acted out the task. Thus, participants did prefer lists but also drew on other formats.

After training participants still preferred a list format (59.4 % of all five HTA after training), not counting participants who combined a list with another format. The second-most frequent choice was a combination of formats (21.1 %), most often a list format combined with a flowchart or another list format. Two new formats emerged: a hierarchy and a narrative (paragraphs of text). Flowcharts and hierarchies accounted for 12.8 % of all trained HTA, whereas narratives and other formats made up the remaining 6.7 % of formats. Thus, participants showed a strong preference for a list format both before and after training.

Training conditions differed in the extent to which participants used lists, flowcharts, or combinations after training ($\chi^2 = 42.23$, $df = 6$, $p < .01$). The majority of participants in the Steps condition preferred to use either lists (45 %) or flowcharts (31.7 %), the latter of which included the two participants who used a hierarchy. In contrast, participants in the Decision-Action Diagram and Concept Map condition had a strong preference for lists (61.7 and 71.7 %) or combined formats (30 and 23.3 %).

Depth and breadth of HTA—The second goal was to describe HTA depth and breadth. If participants created the appropriate procedural knowledge from the training material, then HTA conducted after training should be deeper than before. HTA depth was determined at its deepest level by counting how often a participant created subdivisions. Before training, *making sandwich* had an average depth of 1.3 subgoals (SD = .5) and *making phone call* had a depth of 1.1 (SD = .3). Figure 3a shows two analyses with a depth of one, representative of the initial, untrained HTA (Task 1).

To understand the effect of training, HTA for *making sandwich* and *making phone call* were compared between participants who analyzed these tasks as their Task 1 (before training) and participants who analyzed these tasks as their Task 4 (after training). Figure 4 shows that HTA for Task 4 (after training) were significantly deeper than for the initial Task 1 (*making* *sandwich*: $F(1,36) = 15.85$, $p < .01$, $\eta_p^2 = .35$; *making phone call*: $F(1,36) = 16.81$, $p < .01$, $\eta_p^2 = .36$). Figure 3b shows an HTA produced after training with a depth of three: the first level is labeled "goals", the second level is labeled "subgoals", and the third level is the number list. These data show that training was successful in deepening the HTA.

To explore the nature of HTA breadth, we determined if participants naturally rendered the highest level of their HTA between three to eight subgoals as recommended by literature. For example, Fig. 3a shows a breadth of five (left) and three (right), and b shows a breadth of three. Before training, the average breadth on the highest level for *making sandwich* was 5.5 subgoals (SD = 2.4) and for *making phone call* was 4.2 subgoals (SD = 1.7). HTA breadth did not significantly differ between Task 1 (untrained) and Task 4 after training (*making sandwich*: $p = .64$; *making phone call*: $p = .81$) and was within the desired range of three to eight subgoals, albeit at the narrow end.

A repeated measure ANOVA for breadth and depth of the five trained HTA (by task order) showed that depth and breadth remained the same across the five trained tasks ($p_{\text{depth}} = .12$; $p_{\text{breadth}} = .50$), training conditions ($p_{\text{depth}} = .21$; $p_{\text{breadth}} = .19$), and counterbalance versions ($p_{\text{depth}} = .59$; $p_{\text{breadth}} = .55$). None of the interactions was significant. The above analysis of novices' procedural knowledge showed that HTA depth improved after training and that the breadth was within limits of breadth recommendations. However, as Table 4 shows, some participants continued to create TA that were shallow and/or too narrow or too broad.

Participants' answers to the declarative knowledge test complete this assessment. If participants recognized that a hierarchy was important to HTA (after all, it is part of the name), they should have listed it as one of the five main features of HTA. However, no participant listed "hierarchical" as a main feature of HTA. This may indicate a lack of awareness of the hierarchical nature of HTA, which is consistent with some participants continuing to produce a depth of one.

Subgoals—To assess what subgoals participants identified, two coders coded 2,417 verb–noun pairs with respect to master HTA. Novices sometimes just specified nouns without the verb (e.g., "phone"), which were not coded. We also noticed, but did not quantify, that novices tended to chunk subgoals. For example, one bullet point would have three subgoals listed in one sentence rather than each as a sub-bullet.

The first question was if participants understood the importance of subgoals to HTA. Participants illustrated this in three ways. First, "subgoal" was one of the top-three recalled features in the declarative knowledge test (75 % of participants). Second, participants included the label "subgoal" in 33.9 % of the five task completed after training as seen in Fig. 3b. Third, the total number of subgoals listed for *making sandwich* and *making phone call* doubled after training: Overall, participants identified 233 subgoals when analyzing these as Task 1 (untrained) compared to 473 subgoals noted by participants who analyzed the two tasks after training (Task 4). This represents an increase from an average nine subgoals per participant (*SD* = 4, range: 4–20) to 16 subgoals (*SD* = 6, range: 5–25) for *making sandwich* and an increase from four subgoals (*SD* = 3, range: 0–9) to 10 subgoals

(*SD* = 6, range: 1–24) for *making phone call*. Participants thus indicated on a number of measures that they understood the importance of subgoals to HTA.

If participants generated the required procedural knowledge for redescription, they should include main level subgoals. Participants were expected to identify more main level subgoals for unfamiliar tasks and if they were in the Concept Map condition. Although participants identified more subgoals after training, their focus of analysis remained on lower level subgoals. As Fig. 5 illustrates, participants identified the same proportion of main level to lower level subgoals before and after training for *making sandwich* (p = .72) and for *making phone call* (p = .62). Participants also identified the same proportion of main level subgoals (15.9 %) to lower level subgoals (84.1 %) for general and unfamiliar tasks (p = .82), irrespective of training condition (p = .43). Thus, participants chose a low level of analysis even for tasks for which they did not have specific details.

What subgoals did novices identify? Most subgoals for the cooking tasks were in the "follow recipe" category and the focus of *making sandwich* (89 % of subgoals). For *making breakfast*, participants also emphasized "follow recipe" subgoals (56 %) but devoted some attention to "get recipe" (16 %) and "serve food" (16 %). This stands in contrast to *making Vetkoek* for which participants went into depth for "get recipe" (33 %) compared to *making breakfast* (16 %) and *making sandwich* (4 %). Hardly any subgoals (7 %) were devoted to "enjoy food" and "wrap-up" activities for any of the three tasks.

Verb-noun pairs for tasks in the communication domain were distributed more equally across the categories of the master HTA than for the cooking tasks. One notable exception was the low number of subgoals pertaining to wrap-up activities such as "end call" (9 %), "end meeting" (0 %), and "end sharing" (1 %). *Sharing pictures* had a large number of extra subgoals (29 %) that specified downloading, installing, and learning how to use Adgers.

To summarize, novices focused their cooking HTA on food preparation, and rarely included wrap-up activities for either domain. In contrast to familiar tasks, participants devoted a third of subgoals to preparing for and learning about unfamiliar tasks.

Plan—As plans are an important component of HTA, we assessed whether participants understood this. Participants illustrated in two ways that they recognized the importance of plans to HTA. "Plan" was one of the top three recalled features in the declarative knowledge test (72 % of participants). Participants also used the label "plan" in 36 % of TA completed after training, typically attaching this label to the lowest level of analysis. Although participants recognized the importance of plans to HTA, few participants explicitly devoted space of their TA to it. In addition, participants specified only one plan and implied the hierarchical order of goal—subgoal—plan.

Main goal, criteria, and versatility—Because mentioning the *main goal* and satisfaction *criteria* as well as *versatility* of a TA were dichotomous (yes/no, general/specific), each TA received a composite score. A score of three reflects a "good" TA, containing the main goal, satisfaction criteria, and being general (i.e., included at least three variations).

First, we assessed whether participants' naïve understanding of TA included mentioning of main goal and satisfaction criteria. Based on Patrick et al. (2000) we expected the untrained TA to be specific. Novices' initial TA did not show a good quality. The majority of participants (69 %) scored zero, and as a group, participants only reached an average of 12 % (SD = 20) on the composite score. Only one participant mentioned the main goal, two mentioned satisfaction criteria, and only 10 of the initial TA were general. Thus, novices were not inclined to include the main goal and criteria without instructions and as expected, preferred creating specific TA.

If participants generated the required procedural knowledge from the training material, then their composite scores should be higher after training. As expected, novices who analyzed *making sandwich* and *making phone* after training (Task 4) created significantly better TA (M = 56 %, SD = 31) than novices who analyzed those two tasks before training ($F(1,33)$ = 48.70, $p < .01$). Median and mode of the composite score increased from zero (Task 1) to 67 % (Task 4), that is, two (of three) features. The quality of TA remained at an average of two features over all trained tasks (M = 65 %, SD = 28) and did not significantly differ across trained tasks ($p = .09$) or training conditions ($p = .22$). Although overall performance improved after training, only 28 % of TA had a perfect score, and another 28 % of TA had no or only one feature. Thus, training was successful in improving the quality of TA on the three features, but not for everybody.

What were some typical errors? Participants were least successful at creating general TA. Versatility of the TA did not significantly differ between untrained Task 1 (28 %) and trained Task 4 (42 %, $p = .08$). Of all trained TA, 59 % were general. We expected participants to create more general TA for unfamiliar tasks and if they were in the Concept Map condition. However, a Chi square analysis showed that participants created as many general (or specific) TA for unfamiliar tasks as for familiar tasks ($p = .27$), and the pattern did not significantly differ across training condition ($p = .49$). Thus, unfamiliarity with a task did not prevent participants from producing a specific TA, which shows that creating a general TA was not easy and participants need further instructions.

Although participants were most successful at mentioning the main goal, an error emerged here as well. Seventy one percent of all trained HTAs included the main goal as given to participants (e.g., *making a phone call*), which is in line with 75 % of participants listing the main goal as one of the top three HTA features in the declarative knowledge test. However, about half of the participants "adjusted" the main goal at least once (17 % of all TA), for example by abbreviation to "a good sandwich", or change from "*sharing pictures using Adgers*" to "allow others to see pictures which have been shared with Adgers" and from "*arranging a meeting*" to "have a meeting". Novices' tendency to adjust the wording of the main goal is worthwhile noting because it may lead to an analysis different from the one requested.

Decision factors—To gauge strategies, three questions of the *Task Analysis Questionnaire* at the end of the experiment prompted novices to share their approach to TA: "How did you decide on the *depth* of the analysis, that is, to which level to analyze", "How did you decide on the *breadth* of the analysis, that is, where to start and where to end the task?", and "How did you identify the goals and subgoals?". Two coders segmented and coded participants' responses. Coder agreement was 94 % for depth (Kappa = .92), 90 % for breadth (Kappa = .85), and 97 % for goals/subgoals (Kappa = .96).

Two main strategies emerged, accounting for 92 % of all comments: using a process and using a definition. Process factors included those that referenced a person (e.g., prior knowledge, task familiarity, fatigue), a task (e.g., task complexity), and other, such as asking questions, determining logical order, being specific, being shallow, considering problems, eliminating ambiguities, thinking of the simplest way to do it, or being detailed. Some of the same factors were mentioned as a reason to increase or to decrease analysis depth.

Definitions mostly pertained to that of goals and subgoals and breadth of the analysis. Participants defined a *goal* as "basically the task", "pretty much given", "the big picture", "the main part", and "final product". *Subgoals* then "were the things needed to meet those goals", "each step was a subgoal", and "the elements which were necessary to get the goal, however not broken down into steps like the plan" and "open to my interpretation".

Breadth-related definitions focused on specifying the starting and the ending point. A starting point was "the first step", "whatever step would begin the actual process", "gathering of all relevant information", "the biggest question", and "whatever seemed logically correct as to a beginning". The ending point was "when the tasks were completed", "when the goal was met", or participants "decided not to make it too long" and "stopped before another task would have occurred (prompt was: making breakfast not making and eating)", "to *arrange* the meeting. So it was arranged. Not in participating in it".

Discussion for novices

This study investigated novices' redescriptions before and after receiving one of three types of instructions to inform training of HTA. Table 5 summarizes the desirable outcomes of this training, areas of concerns, and recommendations.

Baseline performance: what to expect from a naïve learner—Data on baseline performance are important to assess the effectiveness of instructions. Where do novices start? Participants preferred to render their analysis in a list-style format and less often as flowcharts. This explains why participants of Patrick et al. (2000) chose these formats even when instructed to use HTA format. However, novices also explored formats such as pictures and motions. Without instructions, TA were shallow with only one (sometimes two) levels deep and rarely contained the main goal. TA were specific to a procedure or technology used and focused on lower-level subgoals, which is consistent with errors that novices make after instruction (Patrick et al. 2000). This suggests that for the naïve learner analyzing a task means unpacking it in some fashion-with one level already providing plenty of detail.

What novices learned—Novices' performance improved on a number of measures after the brief training and practice on five tasks. Participants' declarative knowledge test results showed that the majority of participants recognized that the main goal, subgoals, and plans are important to HTA, which is consistent with Felipe et al. (2010). The HTA themselves were significantly deeper after training, contained about twice as many subgoals, and were of better quality (as defined by mentioning the main goal, satisfaction criteria, and being general). An increased depth, higher number of subgoals and mentioning of the main goal is consistent with findings by Felipe et al. (2010) and shows that novices extracted important aspects of HTA and successfully translated them into procedural knowledge.

Much left to be learned—After training, only few participants did not mention the main goal at all. However, some participants adjusted the main goal as given to them. Such adjustment is not wrong per se and in fact may be a by-product of the overall TA process (Kirwan and Ainsworth 1992). However, it is also important not to change the main goal, once agreement has been achieved. Training could address this topic and increase the quality of HTA. Another error that novices make is not specifying plans on every level of analysis (Patrick et al. 2000; Shepherd 1976). Data from this study suggest that this may be because novices tend to think of a plan in terms of *one* specific way to complete the overall task, associated with the lowest level of analysis rather than every level of analysis.

Despite the success of deeper HTAs, the present data show that the idea of a hierarchy—what it is and looks like—needs further instruction. We suspected that the word "hierarchical" in HTA could be a give away for the declarative knowledge test. However, similar to Felipe et al. (2010), no participant noted that this was a main feature of HTA. Furthermore, some participants continued creating HTA with a depth of one, which is problematic given that HTA depth is a prerequisite to other HTA concepts such as equivalence that we mentioned earlier. Few participants spontaneously used a hierarchy (tree diagram), which is consistent with Felipe et al. (2010). Thus, the HTA diagram itself requires targeted instruction. Given participants strong preference for a list format, training could start with a list, introduce the tree diagram later, and address other formats such as narratives. The differences we found in format choice by training conditions is unclear and may be spurious, given that Felipe et al. (2010) did not report such findings.

Participants indicated that they recognized the importance of subgoals. However, consistent with previous research (Patrick et al. 2000; Stanton and Young 1999), most identified subgoals fell at a lower level. Even without procedural details (unfamiliar tasks), participants preferred to focus on whatever details they knew (e.g., how to obtain a recipe for Vetkoek) rather than outlining higher level subgoals which they should have known. Thus, an unfamiliar task is insufficient to refocus novices' attention to a higher level of analysis.

Somewhat related, novices' HTA tended to be specific, which is consistent with previous findings (Felipe et al. 2010; Patrick et al. 2000). Although the number of general HTAs was higher on the trained task (28 vs. 42 %), this difference was not significant, which can be viewed to support criticism of using simple tasks for teaching. This study suggests that simple tasks have merit, but there are limits to what concepts can be taught by using simple tasks. An alternative explanation is that the idea of a general HTA develops slowly; general HTAs continued to increase to 59 % across the trained tasks.

The differences between training conditions were minimal and provided little support for a differential effect of spatial diagrams. This may be due to the brief amount of training in duration and content, the absence of feedback, or the selection of tasks. Felipe et al. (2010)

found that participants in the Concept Map condition identified significantly more subgoals than Steps and Decision-Action diagram conditions, when analyzing four specific and one unfamiliar task (*making Vetkoek*). In contrast, participants in this study analyzed tasks that varied from specific to general to unfamiliar. The choice of tasks to use for training may be more important than how the information is rendered, at least initially.

To include or not to include—The majority of HTAs were within the recommended breadth range. Yet, participants also created HTAs after training that were too narrow or too broad. This is consistent with a finding by Patrick et al. (2000) indicating that novices have problems determining the correct task boundaries. More specifically, we found that novices tended to forget task completion subgoals for both domains. This blind spot is not trivial and may influence the overall outcome of the TA because potentially important sources of errors are overlooked or new design may not have required functionalities.

Strategies to HTA—Patrick et al. (2000) found that participants used a sequencing or breakdown strategy. Participants in the present study are best described as having used a process or definition. The latter is especially interesting for training purposes, because it points out the necessity to provide clear definitions, for example, for subgoals. However, it also points to the problem that current definitions and differentiation of goals from other concepts (actions, functions) are unclear and a general source of confusion in TA (e.g., Diaper 2004), which has lead some authors to suggest abandoning these concepts altogether (e.g., Diaper and Stanton 2004).

Study 2: Practitioners

Although understanding novices' performance and errors is important for training, studying experienced performers can provide valuable insights. In fact, many TA methods use subject matter experts to understand the knowledge and strategies involved in task performance and inform the design of training (Hoffman and Militello 2009). Just to name a few examples, curricula informed by (cognitive) TA methods have improved learning outcomes in medicine (Luker et al. 2008; Sullivan et al. 2007), mathematical problem-solving skills (Scheiter et al. 2010), as well as biology lab reports and lab attendance rates (Feldon et al. 2010).

Studying experienced performers provides information about the goals of skill development, and the benefits of knowing goals have been shown to be an important factor in training (e.g., Adams 1987). To understand a skill, it is critical to obtain a picture of what experienced performers are actually superior in and to what stimuli and circumstances the skill applies (Ericsson and Smith 1991). This study focused on describing what practitioners would do given similar constraints as novices in Study 1, both to place novices' performance into perspective and to gather (presumably) superior performance. We will refer to this as TA rather than HTA, given that participants may not have intended to use HTA. Our focus remains on subgoal redescription.

Two possible approaches for redescribing subgoals are differentiated based on whether a task analyst chooses to analyze the breadth or the depth of a task first (Jonassen et al. 1999). A breadth-first approach means to redescribe the main goal into lower levels before moving on to the next level. Using numbers indicating levels such as those shown in Fig. 1, the sequence of subgoals identified might look like this: 1.0, 2.0, 3.0, 4.0, 5.0, 1.1. 1.2, 1.3, 1.4, 2.1, 2.2, 2.3 and so forth. Visually this might look like a line that undulates horizontally. Conversely, an analyst using a depth-first approach will start redescribing the main goal into the first subgoal and continue to move down and up in depth, in effect showing a sequence such as 1.0, 1.1. 1.1.1, 1.1.2, 1.1.3, 1.2, 1.3, 1.4, 2.0, 2.1., and so forth. Visually this approach might look like a line that undulates vertically.

Another strategy is to ask questions, often in the context of eliciting knowledge (e.g., Stanton 2006). Two general questions guide the instructional designer during the principled skill decomposition phase. "Which skills are necessary in order to be *able* to perform the skill under investigation" (van Merriënboer 1997, p. 86) is meant to elicit elements on a lower level in the hierarchy, and "Are there any *other* skills necessary to be able to perform the skill under consideration" (p. 87) helps elicit elements on the same level in the hierarchy. Stanton (2006) compared different lists of specific questions that varied based on the problem domain a task analyst is working in. We chose six general questions (what, when, where, who, why, how) to investigate in more depth how questions guide a practitioner during redescription.

Assumptions can be viewed as the flip-side of questions, namely when analysis has to progress but there is nobody to answer questions. Furthermore, stating assumptions is an important part of the analysis because it helps understand the limitations and applicability of the analysis (Kieras 2004). Thus, we assessed whether experienced task analysts did indeed make assumptions, and if yes, what those were.

Overview of Study 2

To summarize, the goals of this study were to determine the characteristics of TA products of experienced practitioners along some of the same criteria as in Study 1. In addition, we wanted to characterize practitioners' approach. To gather information about characteristics of experienced task analysts' products and process, participants in Study 2 analyzed six tasks while thinking aloud, completed questionnaires, and participated in a semi-structured interview (the data for which are not presented here).

Method for practitioners

Participants—Four of the eight practitioners participated in Atlanta (GA) and four participated in Raleigh (NC). All participants (2 male, 6 female) spoke English as their native Language (see Table 6 for participant characteristics). Most of the participants were Caucasian (5). Six participants indicated a master's degree as their highest level of education, one had a doctorate. The majors were Industrial Engineering, Biomedical Engineering, Industrial Engineering, Instructional Design, Rehabilitation Counseling, Occupational Ergonomics, and Psychology. Licenses included Certified Professional Ergonomists, Industrial-Professional Engineer, and Occupational Therapist. The study lasted approximately 3 h for which participants received a $50.00 honorarium.

Recruitment—Participants were recruited via professional organizations and companies whose members were known or likely to use TA: Human Factors and Ergonomics Society, Special Interest Group on Computer–Human Interaction, Instructional Technology Forum, and the Board of Certification in Professional Ergonomics.

Task analysis experience and self-rated proficiency—To be included in the study, participants needed to be native English speakers, use TA in their job, have at least 2 years experience conducting TA, and worked on at least one TA in the past year. Two years experience should ensure that participants experienced some breadth in their TA work without having advanced to a managerial position.

Participants expressed a range of experience with TA as assessed by the *Demographics and Experience Questionnaire* and shown in Table 6. In the past year, six participants had conducted relatively few (2–5) TA whereas two participants had conducted many (30–50). Over the course of their professional life, two participants had conducted fewer than 5 TA, one participant conducted between 6 and 12, and the remaining 5 participants indicated that they conducted more than 50 TA. TA methods that participants reported using reflected the variety of existing methods. One participant stated using every type depending on the circumstances, whereas another participant did not know the formal names of the methods used. Participants reported learning the methods on the job (46 %), in school (43 %), or in a course (11 %). As for specific TA methods that we queried participants about: five participants had heard about CTA, three about HTA, and two about SGT.

TA are undertaken for a particular purpose and have specific, measurable goals. Participants' top three *purposes* for conducting TA were designing tasks; designing equipment and products; and training individuals. Less frequently mentioned was environmental design. No participant used TA to select individuals, but they did for identifying barriers to person-environment fit and selecting jobs for individuals with disabilities. The top two *goals* for conducting TA were to enhance performance and increase safety. Increasing comfort and user satisfaction was also a goal for half of the participants, but less frequently. Only one participant used TA to find an assistive technology fit for a person.

The tasks that participants analyzed in their work were diverse and spanned from household work to repairing an airplane. Tasks were those found in military, repair and vehicle manufacturing, factory, office, work, and service industry environments. More specific descriptions included graph construction, software installation, and authentication. Participants also listed complex performance (equipment diagnostics, equipment operation), cognitive tasks (decision-making, critical thinking), aircraft maintenance, as well as various airport and airline tasks. Participants moreover reportedly analyzed how a person works at a desk, performs various household activities (e.g., cooking or cleaning), specific computer tasks, uses a telephone, or checks in at a hotel.

Instructions for task analysis—Participants received a scenario that described them joining a new team. The new team members had asked the participant to create common ground by illustrating her/his understanding of TA on a number of example tasks. To capture participants' approach, we neither provided a purpose for conducting their TA nor instructed participants to focus on a specific TA phase/method.

Tasks to be analyzed—Practitioners analyzed the same tasks as novices in Study 1.

Questionnaires—Over the course of the study, participants completed the same three questionnaires as novices had in Study 1. The *Demographics and Experience Questionnaire* also probed for information about certifications, experience with TA, for what purposes and goals participants used TA, and what aspects of a task participants emphasized in their analysis. Participants listed the TA methods they used and indicated how often they used them, when and how they learned them, and rated their own proficiency. Then questions specifically targeted experience with five TA methods, including HTA and CTA. In the *Task Analysis Questionnaire*, participants also rated how representative their TA was in comparison to the ones in their job.

Equipment and set-up—An Olympus DM-10 voice recorder taped all interviews. Participants conducted their TA on 11 × 17 in. paper, placed in landscape format in front of them. Two QuickCam web cameras (Logitech 2007) and Morae Recorder software (TechSmith 2009) captured participants' hands and workspace from two different angles while participants completed the TA.

Design and procedure—As in Study 1, this study incorporated repeated measures as participants analyzed six tasks, arranged in two counterbalanced orders. Participants read and signed the informed consent form. Then the experimenter collected the *Demographics and Experience Questionnaire* that had been mailed to participants prior to the study. Participants were oriented to what the video cameras captured before the video recording began. Familiarization with thinking aloud and being recorded occurred by playing tic-tac-toe with the experimenter. Then, participants read the scenario that asked them to illustrate their understanding of TA on a number of example tasks to the new team they joined. For each task, participants received a written task to be analyzed (as shown in Table 1) and instructed to perform the TA while thinking aloud. The experimenter collected the TA and provided the next task when participants indicated that the TA was complete (putting down the pencil), latest after 15 min. After three TA, participants took a 5-min break. Once all tasks were analyzed, participants completed the *Task Questionnaire* and *Task Analysis Questionnaire*. Participants took a 10-min break before beginning the semi-structured interview (data are not presented here) after which they were debriefed.

Results for practitioners

Data analysis focused on two areas. First, participants' TA products were examined to determine product characteristics. The TA were coded on the same dimensions as used for novices from Study 1: format of TA, dimensions of the hierarchy, subgoals, quality (main goal, satisfaction criteria, and versatility). Two coders coded all TA with respect to master TA, and disagreements were resolved through discussion. Mean overall coder agreement was 82 % (range 75–86 %, Kappa = .69–.84). Second, participants' think-aloud protocols were analyzed for process characteristics (breadth or depth-first, questions, and assumptions). Coder agreement was 86 % (Kappa = .83).

Format of task analyses—The first question was what format practitioners would choose to render their TA and how prominent a hierarchy featured. We expected TA to reflect the diversity of formats reported by Ainsworth and Marshall's (1998): decomposition tables, subjective reports, flow charts, timelines, verbal transcripts, workload assessment graphs, HTA diagrams, and decision-action diagrams. The majority of practitioners used a list format such as a numbered or bulleted list with indents to indicate different levels of analysis (83 % of TA). Participants also used a flowchart (9 %) or combined formats (4 %). No participant used a hierarchy to illustrate TA, and two TA (4 %) showed a loose collection of subgoals. Irrespective of format, each verb-noun pair was visually separated from other verb-noun pairs in some fashion.

Figure 6a shows two TA examples from an instructional design perspective. Notable here is the decomposition (classification) of subgoals into knowledge, motor skills, and attitude (KSA). This participant was the only one who had a similar task structure for all three cooking tasks. Other participants would create a similar TA but without the KSA classification. A TA from a system design perspective is shown in Fig. 6b. Although not formally a hierarchy (tree diagram), the levels of analysis of goal, subgoals, and actions are clearly labeled. Also documented are the assumptions on the right.

Figure 6c illustrates a TA focused on assessing task performance of a patient with brain injury. This practitioner first documented assumptions about the patient (usually a given) and then outlined what to assess (e.g., cueing, sequencing, problem solving, judgment). This practitioner would then ask the patient to go through task steps while checking/evaluating if the performance was adequate. Thus, our sample did not reflect the diversity of formats found by Ainsworth and Marshall's (1998). Although some participants indicated different levels of analysis or using HTA and GDTA as methods they used in their work, nobody used a hierarchy during the 15 min of illustration.

Depth and breadth of task analysis—Practitioners' TA were expected to be at least two levels deep, reflecting different levels of analysis, and they were on average 2.3 levels deep (SD = .95), ranging in depth from one to six levels. A non-parametric Friedman test showed that TA did not differ significantly in depth across tasks ($p = .88$). Thus, practitioners created TA of more than one level for specific tasks such as *making sandwich* or *making phone call*, as well as when no specific details were available (unfamiliar tasks). However, as shown in Table 4, some participants created TA that were only one level deep. As expected most, but not all, practitioners redescribed their TA.

Breadth was expected to be within suggested boundaries of three to eight subgoals derived from literature as outlined earlier. TA breadth was on average 6.1 subgoals wide (SD = 4.23), ranging from 2 to 21 subgoals. A Friedman test showed that tasks signifi-cantly differed in their breadth ($\chi^2 = 11.67$, $df = 5$, $p = .04$); however, follow-up multiple comparisons using the Wilcoxon test and a Bonferroni-adjusted alpha-level did not indicate significant differences between all pairs. Although the average breadth of practitioners' TA was within the suggested boundaries of three to eight subgoals, Fig. 7 shows that practitioners in this study created TA that were beyond those boundaries. The broadest and shallowest analyses were created for *making sandwich*, with two participants creating the broadest TA of 19 and 21 elements. This illustrates practitioners' individual differences and that they do not necessarily adhere to the breadth standards suggested in the literature.

Subgoals—Of particular interest was what subgoals practitioners included and excluded from the TA. In general, practitioners were rather specific in their analysis with 5 % of all identified subgoals matching to a main level subgoal of our master, 90 % of subgoals focusing on lower level subgoals and 5 % were extra. As Fig. 8 shows, practitioners mentioned on average only one main level subgoal from our master TA and identified almost three times as many subgoals for cooking tasks compared to communication tasks. Because novices presumably have inappropriate task boundaries (Patrick et al. 2000), we will now describe for each task what subgoals practitioners included.

Subgoals for cooking tasks—More specifically, for the task of *making sandwich*, participants concentrated on describing the procedure (80 % of 176 subgoals), with rarely mentioning to determine what to make (part of get recipe) or serving the sandwich. However, five of eight participants included the main level goal of enjoying the sandwich. A notable number of the verb–noun pairs (13 %) were devoted to wrapping up, that is, cleaning.

For the task of *making breakfast*, participants also paid most attention to the preparation of breakfast items (64 % of 241 subgoals), focusing on food and rarely mentioned beverage. However, some verb–noun pairs were devoted to "determining what to make" (13 %), which is not that surprising, given the task of making breakfast includes choices that are already made for the task of *making sandwich*. Another 10 % of subgoals were devoted to serve breakfast. And again, participants noted wrap-up activities such as cleaning the dishes (7 %). Two participants included subgoals such as leaving the room and turning off the lights, which were outside our master task list and coded as extra.

When practitioners analyzed *making Vetkoek* they spent much of their focus on learning what Vetkoek is (39 % of 146 subgoals). These subgoals (get recipe) included determining what Vetkoek is, where it comes from, what ingredients it uses, how to make it, and if they had the equipment and knew the techniques involved in making the dish. Only 43 % of subgoals related to following the recipe, and only two participants noted to enjoy the dish. Again, TA included some wrap-up activities (9 %), which suggests that participants perceived cleaning and storing items as part of the general cooking task structure.

Subgoals for communication tasks—For the task of making a *phone call*, participants focused on subgoals related to determining the receiver (38 % of 65 subgoals), connecting (40 %), and somewhat on communication (15 %). Little emphasis was placed on obtaining a phone (2 %) or ending the call (5 %).

When analyzing *arranging meeting*, participants emphasized determining date and time (28 % of 89 subgoals), determine attendees (17 %), and determine location (17 %). Not so much focus was placed on determining the reason for the meeting (8 %) or confirming the meeting details (7 %). Practitioners invested 17 % of their subgoals to prepare for the meeting, but only 5 % to meet, and none to end and wrap up the meeting. One could argue that the task of arranging a meeting does not include the meeting itself and that this finding should not be surprising. However, one may counter that making a phone call does not include the conversation either; yet, participants included it in their TA. Nobody included any items related to ending the meeting.

Last, for the unfamiliar task of *sharing pictures* using Adgers, participants mainly analyzed the exchange aspect of making the picture available (39 % of 51 subgoals), followed by connecting using Adgers (22 %), obtaining the picture (16 %), and determining which picture to be shared (16 %). Only few subgoals pertained to determine receiver information (8 %), and no participant mentioned any subgoals related to end the sharing. However, participants identified an additional 24 subgoals to include efforts to obtain a copy of the software, install it, use a tutorial, and explore the software to become familiar with it, thus including tasks in the TA that they would be doing themselves because of their unfamiliarity with Adgers. Similar to the task of *arranging meeting*, participants did not address the end of sharing pictures as a closing symmetry.

Only one participant pondered about the task boundaries and decided *not* to include learning about the unfamiliar task in the task analysis itself: “I’m trying to decide where I would start since I don’t have a clue what Adgers is. So I’m trying to decide if I would include something like learn what Adgers is, is part of the task analysis. Presumably if I’m doing a task analysis though, I wouldn’t, normally I wouldn’t include something like that, as part of the task of actually sharing the pictures”. This suggests that practitioners who are conducting TA can use their inexperience with a task as a guide.

Subgoal symmetry—As mentioned above, participants’ TA of the cooking domain included symmetrical wrap-up activities such as cleaning and storing away times. There was a noticeable symmetry even on lower levels of analysis. For example, “open jar” was followed by “close jar”, “open the fridge” was followed by “close the fridge”, and “open the sandwich” was followed by “close the sandwich”. Cleaning can be viewed as being symmetrical to the whole sandwich making activity. Participants also included wrap-up activities for *making phone call* (end call), but not for the other communication tasks. Thus, practitioners’ TA contained symmetry but it was not pervasive across all tasks (as we defined it).

Qualities of a good task analysis—Mentioning the main goal, satisfaction criteria, and versatility were assessed for practitioners in the same fashion as for novices in Study 1, that is, they sum up to three for each TA and person and indicate a “good” TA. Overall, the quality of practitioners’ TA on these three categories was 28 %. Practitioners rarely stated satisfaction criteria, but that is not too surprising, given that these criteria are not necessarily part of all TA methods. Surprising was, however, that only 27 % of TA contained the main goal (as given or adjusted), with one participant (instructional design) accounting for half of those.

Only 56 % of participants' TA were general, with one participant creating specific TA for all tasks and another creating general TA for all tasks. Think-aloud data may explain how and why participants created specific TA for unfamiliar tasks (*making Vetkoek*, *sharing pictures* using Adgers). One reason was that participants constrained the problem space very tightly. For example, one participant constrained the TA of *making Vetkoek* so that it only included finding a recipe for Vetkoek in a cookbook. Another reason was being guided by existing technology. For example, some participants thought of Facebook when analyzing *sharing pictures*, and let this knowledge and experience be their guide.

Another explanation for why participants created specific TA relates to the purpose of conducting a TA. To illustrate, one participant who used TA to evaluate the capabilities of a specific person to perform a certain job was thus working with clearly defined parameters. The person whose performance is assessed has very specific capabilities and limitations (e.g., due to injury), performance was evaluated in a very specific environment (e.g., kitchen) and was tied to very specific objects (e.g., phone model). Thus, the resulting TA (assessment) was specific. This stands in contrast to another participant who started out with a particular scenario and then tested how the TA held up when expanding the assumptions to different scenarios, thus creating a general TA needed for system or training design. There are a number of reasons why a TA might be specific, perhaps by design or inadvertently.

Task analysis process—The think-aloud data provided the basis for analyzing the TA process as an account of how participants conducted the TA: (1) Do participants determine first the breadth of the analysis or analyze subgoals in depth first before determining the next subgoal? and (2) What questions do participants ask, and what assumptions do they make? Overall coder agreement between two coders was 86 % (Kappa = .83).

Breadth-first versus depth-first—TA were coded as to whether a participant approached it breadth-first or depth-first. If a participant outlined all subgoals on the highest level first before redescribing these into lower level subgoals, then this was coded as breadth-first, even the TA consisted only of one level. If a participant redescribed subgoals before having outlined *all* high-level subgoals, then this was coded as depth-first.

One participant's comments shed light onto the benefits of a breadth-first approach: "I would start with the breadth-first analysis, 'cause […] what I want to understand is, do I understand the end problem? You know, are there any big gaps in my knowledge about where the user is going to start and where the user is gonna end up?". Besides determining the boundaries of the task, a breadth-first approach also prevents the team from wasting time outlining details of a branch that may be cut out of the project sometime later. Furthermore, having specific details may be counterproductive to creating a shared understanding because software developers may be inclined to code too early in the process.

Twenty one TA (44 %) were created by a breadth-first approach and 27 (56 %) were done depth-first. Cooking tasks were more likely conducted depth-first and communication tasks were more likely conducted breadth-first ($\chi^2 = 5.76$, $df = 1$, $p < .01$). One participant changed from breadth-first to a depth-first approach when moving from the communication to the cooking tasks. This participant explicitly noted this change in approach while analyzing *making breakfast*: "I just realized that I rushed right into the making the peanut butter jelly sandwich without clarifying the assumptions that I had there, which was that the sandwich was for me". This suggests that a breadth-first approach has practical benefits but that procedural/sequential aspects of a task (domain) may influence redescription.

Practitioners' questions during a task analysis—The next goal was to understand what questions practitioners used during their tasks analysis and what these accomplished. Think-aloud data were coded for whether participants mentioned the questions "what, when, where, who, why, and how" during their TA. A segment was defined as an idea unit, containing a question that furthered the TA (i.e., not including questions to the experimenter). The think-aloud protocols were conservatively coded, that is, excluding questions that were phrased as statements. One coder selected the 226 segments and two coders coded them. The coding scheme included an "other" category for questions other than the ones previously mentioned.

All participants asked questions at some point during their TA and varied in the number of questions they asked, from none to 16 for one task and between one and 51 for all six tasks. Most questions in this phase of analysis pertained to *what* (43 % of 226), followed by questions about *how* (16 %). The remaining four questions accounted only for 16 % of the remaining segments, whereas 25 % of the questions were not captured by the six questions in the coding scheme. We will now show some examples in more detail.

What?: Common themes of *what* questions emerged. One category of questions were about understanding the task space and specifying its objects, such as *What is it?*, *What type of jelly? What type of phone? What type of materials? What would you use?*. This was seen especially, but not only, with the two unfamiliar tasks (e.g., *What are the system requirements?*, *What are the capabilities?*). Another category of *what* questions related to the procedure of the task, such as *What is next?*, *What is the process?*, *What are the steps?*, *What is the first step?* and *What do I do?*.

A third category of *what* questions related to searching for specific requirements (e.g., *What would I need? What utensils do I need?*). Other *what* questions included checking specific aspects of a task: *What will I need to know?*, *What will I need to be able to do?*, *What would the knowledge behavior be?*, *What behavior would I use?*, and *What kind of motor skills are involved?*. Participants occasionally asked *What if?* questions to understand alternative paths and asked *What else?* to search the task space for potentially undiscovered task elements.

Questions were also phrased so they fell in a different question category. For example, a *who* question was phrased as *What's the audience?* and a *How long?* question was phrased as *What takes the longest?*. This suggests that participants would rephrase open ended questions into a specific question that would guide them to the next step, in this case, to start with the item that takes the longest. Thus, *what* questions aimed to elicit and specify task objects, information requirements, and procedural details.

How?: Questions that contained *how* were the second most frequently mentioned category (16 % of the segments). Questions mostly related to "*how to*" followed by a verb, for instance, *how to share, how to have, how to use, how to dial, how to open, or how to choose*. These questions suggest the search for a procedure. There were also questions related to number (*how many*), time (*how far back*), assessment of ability (e.g., *how able is he to maneuver?*) and even looking for answers (e.g., *how can he cue himself?*), which suggests that *how* questions can also be used for quantification and evaluation.

When, where, who, and why?: Only 16 % of segments fell into the remaining four question categories. *Where* questions could refer to a location of an object (e.g., *Where is the peanut butter?*) or a starting point of the TA (*Where do I start?*). *Who* questions focused on defining an audience, both more general as in *Who do you want to share the pictures with?* and more specifically *Who in the family?*. One participant contributed to seven of the nine *why* questions by questioning the main goal at the beginning of the TA: *Why are we sharing pictures*? or *Why are we using Adgers*?

Other questions: One quarter of the segments included other questions. Participants asked focused questions that required a yes/no answer, for example when assessing behavior (e.g., *Is he able to do x?* or *Is he doing y?*), determining timing (*Are there things that are going to be done in parallel?*), and checking aspects (e.g., *Is there anything I need to know?*) or asked "Do I…" (e.g., *Do I create the agenda before?* or *Do we have everything necessary?*). Furthermore, participants searched for the right word (*stove…use it? Employ it?*) These questions show while analyzing a task, participants also checked the presence of specific task aspects and tried to determine the right labels for the subgoal they had in mind.

Practitioners' assumptions during a task analysis—We expected participants to make assumptions when not enough information was available, especially at the beginning of the analysis given our general instructions. Think-aloud protocols were inspected for whether participants mentioned assumptions, which were conservatively defined as an idea unit containing the words "assume", "assuming", or "assumption". Overall, participants stated 69 assumptions, 80 % of which came from two participants. The following categories of assumptions emerged: assumptions about the user (e.g., who am I making this for), experience (e.g., I have used/never used this before), ability (assume that he can/cannot reach), location and prerequisites (e.g., "I assume I have a kitchen and the ingredients are already there so I don't have to go out and buy them"), particular make up of an object (e.g., assume a jar—as opposed to other peanut butter containers). Data showed that participants also rejected assumptions once they were made. Also, as one participant pointed out "So end user…is me. And there's assumptions embedded in what me means". Data showed the participants made (and rejected) assumptions in the beginning, but that the TA analysis itself also brought about decision points where assumptions were necessary to continue.

Discussion for practitioners

The goal of Study 2 was to capture characteristics about the products and processes of experienced task analysts. Collecting information about skill expression is one step in studying expertise (Ericsson and Smith 1991) and provides information about goals of skill development and some basis against which to adjust current performance.

General characteristics of practitioners' task analysis product—In 15 min, practitioners created a TA that was most often rendered in a list format with each individual subgoal placed separately. Most of the participants focused on redescription, with only one practitioner illustrating how to evaluate performance based on subgoals. TA had an average breadth of six subgoals on the highest level and an average depth of two to three levels. Depth was independent of task, which suggests that participants have a certain depth in mind for their initial draft, even for unfamiliar tasks, that is, when specific details were unknown. These data provide ballpark numbers for novice products.

Stating the main goal provides important context for an analysis, and we expected participants to do this irrespective of the purpose of the analysis. We were surprised that one participant (instructional design) accounted for half the TA that contained the main goal. Thus, the presence of a main goal is not a good predictor of redescription quality—at least not in the first draft and given the limitation that some participants were less experienced than others.

Deviation, draft, or inferior product?—The majority of the TA (60 %) fell within the suggested range of three to eight subgoals; however, some TA were as small as two and as broad as 21 subgoals on the highest level. Breadth varied between tasks, but not as a function of familiarity. This illustrates that although practitioners' TA were mostly within the suggested breadth boundaries, they did not necessarily adhere to the standards suggested in the literature, at least not for an initial draft created within 15 min. Participants also created TA of a depth one, which either illustrates suboptimal performance or suggests that this level of analysis is all that is required for some practitioners (e.g., to assess performance).

The importance of defining the purpose of conducting a TA is mentioned throughout literature along with the emphasis that it influences the TA (e.g., Kirwan and Ainsworth 1992). Participants in this sample did not receive a specific purpose yet focused on redescription, except for one practitioner who did not create but rather used a redescription for assessment. Furthermore, half of the TA were considered specific to a person, technology, or procedure. One advantage of a goals hierarchy is that it is generalizable and technology independent—at least on the higher levels of analysis (e.g., Annett 2004; Endsley et al. 2003). Think-aloud data suggested that versatility depended on the purpose of the TA, along with how tightly participants constrained the task space and whether they used another technology as a guide.

Figure 9 illustrates how a system designer may be concerned with how a variety of people will make a phone call using a range of phones (e.g., cell phone or landline). Hence, the TA needs to consider different person and phone variables, which means accommodating a number of scenarios. In contrast, an Occupational Therapist's or Ergonomist's concern may be one particular person with a unique combination of abilities or injuries. The focus is then on whether this person can accomplish, for example, the goal of dialing the number on one particular phone, thus considering one point in the task space at any given time. This is important to keep in mind when assessing novice performance because the same outcome can have different reasons: A practitioner may explicitly choose a level of versatility whereas a novice may do so by default.

Subgoals & task boundaries—Compared to our master TA, practitioners focused on identifying lower level subgoals. Only 5 % of the subgoals were identified on the highest level in the master TA (one average one subgoal) and another 5 % were extra, that is, outside of the boundaries of the master TA. These patterns are beneficial to understand novice performance given the same task constraints. Participants also included subgoals such as learning about unfamiliar tasks; however, one participant reflected on this notion and decided to exclude it. Unfamiliarity with a task is said to be an advantage for the task analyst (Shepherd 2001); yet these data suggest that task unfamiliarity may also be a pitfall for drafting a TA before meeting with subject matter experts.

Participants included subgoals for tasks in the cooking domain and for *making phone call* that can be described as symmetrical or complementary. For example, if there was a subgoal that specified to open a jar, drawer, or fridge, then there was another subgoal that specified to close it. The importance of considering symmetrical subgoals is evident when considering a storage system (a warehouse, a data disk, or working memory) that only allows to "*add items*" but not to "*delete items*". In the case of *sharing pictures* using Adgers, participants could have used this as an important cue to check if the analysis was complete and notice that they uploaded a picture but that, according to their draft, it lived on the web forever.

A dynamic process—When reading introductions to TA, the iterative nature of the process is emphasized, yet instructions appear static (one-directional) and intermediary products not displayed (e.g., Shepherd 2001). Flowcharts illustrating how redescription occurs make it appear as if a hierarchy is being created top–down, with questions designed to guide how each level is fleshed out before moving on to the next level. This reflects less the act of "creating" a goals hierarchy and resembles more the act of retrieving a well-formed representation, such as one that an analyst may learn with repeated analyses of the same domain (Shepherd).

Participants in this study illustrated that creating the redescription is a dynamic process. Many questions that participants asked are similar to those compared by Stanton (2006). However, participants used these questions to understand the task space, narrow or broaden it, search it for objects and requirements, elicit information about procedure. They negotiated which subgoals to include, how to name them, and where their place should be, even switching them around. This resembles the qualitative data analysis approach that (Cognitive)TA uses to find meaningful findings (Crandall et al. 2006)—only that the unit of analysis in redescription is a verb–noun pair rather than cognition.

In GDTA, the practitioner can expect the initial draft of a goals hierarchy to change when reviewing the draft with the subject matter expert by adding, deleted, or rearranging goals (Endsley et al. 2003). Our data show that this can also happen when constructing the initial draft. Endsley et al. (2003) differentiate goals (e.g., *provide effective communication*) from tasks (e.g., *make radio communication*, technology dependent and to be physically accomplished) and information requirements. The latter two provide important cues as goals can be derived by then asking "why" and "what for". This may feed into efforts of automating task analysis (e.g., Asimakopoulos et al. 2011) that parse specific actions from procedural scenarios in an attempt to infer task structure.

It appears then that creating a goals hierarchy can be described as a three-stage process. First, it is outlined "what" is done, then goals are revised by asking "why", and Endsley et al. (2003) also recommended asking "why" to position a subgoal in the goals hierarchy. Asking "why" is also a major component of the approach to extract subgoals as described by Catrambone (2011). "Why" questions were rarely asked by our participants, which could suggest that they were still in the first phase and we could expect some revisions.

Overall discussion

Although the process and methods of TA have been described and skills identified (e.g., Hoffman and Militello 2009), expertise in TA has yet to be defined. It has been suggested that TA is the process of understanding a task and associated knowledge or problems, and that its value lies in creating the shared understanding of the TA team (e.g., Shepherd 1976). However, a product such as a redescription is just as valuable, given that it feeds into subsequent analyses (e.g., SHERPA, GOMS) and could affect subsequent results. We therefore operationalized expertise in (H)TA to consist of both the process and the product (HTA diagram) and compared novice and experienced performance on the same set of tasks using the same master TA (coding scheme).

The goal of this research was to determine skill components in redescription by characterizing novices' and practitioners' performance on the same set of tasks. These studies captured a first draft of a redescription, completed within 15 min, based on a main goal given to participants. Table 7 shows a summary of the skill components and guidelines for training. This table is by no means complete given that (H)TA is an iterative process and includes many other activities and skills such as interviewing (Ainsworth 2001) and bootstrapping (quickly learning about a new domain; Crandall et al. 2006). Thus, it remains unanswered how this initial draft may change given more information, thought, and analysis.

What to expect from a novice

Novices (after training) and practitioners performed similarly on some of the measures. For example, their redescriptions had similar breadth and depth dimensions, and both groups created HTA with a depth of one and had breadths that were outside of breadth recommendations. This suggests that (a) the dimensions of novices' HTA reflect what one can reasonably expect from a first draft, and (b) that a depth of one may be enough to be useful for some practitioners but that for novice this indicates suboptimal performance. Knowing what to expect from a novice is important for a teacher to gauge performance, especially someone who has little practical HTA experience.

Tasks

On average, novices and practitioners both identified more subgoals for the cooking domain compared to the communication domain; however, for practitioners the difference was threefold. An obvious explanation is that practitioners tended to be older and thus probably had more cooking experience. However, given that both groups have experience in the communication domain, we interpret this difference to reflect practitioners' focus and appreciation of task details (verb–noun pairs) that are more easily accessible and manifold when analyzing some tasks. Tasks such as the making a peanut-butter jelly sandwich may be perceived as being too simple to learn TA. Yet, our data suggest that cooking tasks are well suited to convey the idea of attention to detail, thus leaving complex tasks for training of more advanced concepts.

Unit of analysis

Table 7 is a summary of components based on findings from both studies. To illustrate how we arrived at them we will consider an area in which practitioners and novices differed: subgoals. For example, practitioners clearly delineated verb–noun pairs by placing individual ones on a different line or bullet, which indicates a small chunk size. Novices, however, often clumped three or four verb–noun pairs into one bullet without any subdivisions. One difference between novices and experts (or highly skilled performers) is how they represented a problem, in this case a task (e.g., Ericsson and Smith 1991). Our data suggest that for redescription, this means that a problem (the task) is presented in small units (one verb–noun pair) that are chunked together. This also means that the size of the unit of analysis might predict TA experience.

Using this and other findings we can unpack instructions on redescription that say "state subordinate goals" (e.g., Stanton 2006). To state subgoals then means that one has to define what a subgoal is, identify the subgoal, delineate it from other subgoals, determine how to label it, and notice if it outside the boundaries. We suggest then that training for novices start with recognizing and separating the unit of analysis (verb–noun pairs), how they are or can be arranged in the hierarchy before advancing to other concepts such as determining whether levels of analysis are equivalent to each other.

Measures of quality

The quest for how to determine the reliability, validity, efficiency, and effectiveness of a TA is ongoing, and it has been suggested to understand qualities of poor and good TA instead (Hoffman and Militello 2009). Some have argued that a good TA is one that produces useful results (e.g., Shepherd 1976) and captures task aspects one intends to measure, for example, cognition, and is grounded in theory (Crandall et al. 2006). However, these definitions may lead to devalue a good analysis when there is just not much to be found (Ainsworth and Marshall 1998).

Measures to determine the quality of an HTA diagram include being hierarchical, equivalence between levels of analysis, logical decomposition, and versatility of the analysis (Patrick et al. 2000). Some of these qualities depend on the existence of others. For example, a hierarchy is a prerequisite to determine if different levels of analysis are equivalent. We combined recommendations from literature to create different operational definitions and assess task redescriptions. The goal was to show that these measures are useful in describing novice and practitioner performance. However, they also have limitations. For example, whether a redescription is versatile appears to depend on the TA purpose and thus should be considered during assessment of a redescription.

Novices desire validation of their redescriptions, but this request has been countered with the argument that there are many valid ways to redescribe a task (Shepherd 1976). Thus, the description of a hierarchy in terms of breadth and depth may come with the caveat that it only provides a relative assessment rather than an absolute one. Nevertheless, recommendations for breadth and depth are informed by boundaries of what analysts can conceptualize (Ainsworth 2001), experience teaches that some redescriptions are more useful than others (Shepherd). Thus, quantifying the dimensions of a redescription may be beneficial after all for an initial, visual assessment.

Conclusion

This research suggests in more detail how redescription occurs based on novice and practitioner data derived from the same six tasks and coding scheme. Crandall et al. (2006) called for new methods and more competent practitioners indicate an increased training need, not only in how to analyze qualitative data but also how to draw a task structure from those data. These skill components address some of the related training questions.

Acknowledgments

This research was supported in part by contributions from Deere & Company. We thank Jerry Duncan for his support and advice on this research. This research was submitted in partial fulfillment of the Doctor of Philosophy degree at Georgia Institute of Technology (Adams, 2010). We thank the many people who helped connect us with practitioners for Study 2, and the members of the Human Factors and Aging Laboratory helped with data collection and analysis. We especially appreciate the many hours of coding task analysis by Sarah (Felipe) Gobrogge and the space provided by Dr. Anne McLaughlin to meet with participants. Lastly, we thank the reviewers for their time and feedback: Drs. R. R. Hoffman & T. Ormerod and one anonymous reviewer.

References

Adams JA. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin. 1987; 101(1):41–74.

Ainsworth, L. Task analysis. In: Noyes, J., Bransby, M., editors. People in control. London, UK: Institution of Electrical Engineers; 2001. p. 117-132.

Ainsworth L, Marshall E. Issues of quality and practicability in task analysis: Preliminary results from two surveys. Ergonomics. 1998; 41(11):1607–1617. DOI: 10.1080/001401398186090 [PubMed: 9819577]

Annett, J. Hierarchical task analysis. In: Diaper, D., Stanton, NA., editors. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004. p. 67-82.

Annett J, Duncan KD. Task analysis and training design. Occupational Psychology. 1967; 41(4):211–221.

Asimakopoulos S, Dix A, Fildes R. Using hierarchical task decomposition as a grammar to map actions in context: Application to forecasting systems in supply chain planning. International Journal of Human-Computer Studies. 2011; 69:234–250.

Carlson R, Chandler P, Sweller J. Learning and understanding science instructional material. Journal of Educational Psychology. 2003; 95(3):629–640. DOI: 10.1037/0022-0663.95.3.629

Catrambone, R. Paper presented at the 2011 Learning and Technology Symposium. Columbus, GA: 2011. Task analysis by problem solving (TAPS): Uncovering expert knowledge to develop high-quality instructional materials and training.

Clark, RE., Feldon, DF., van Merriënboer, JJG., Yates, KE., Early, S. Cognitive task analysis. In: Spector, JM.Merrill, MD.van Merriënboer, JJG., Driscoll, MP., editors. Handbook of research on educational communications and technology. 3rd. Mahwah, NJ: Lawrence Erlbaum Associates; 2008.

Craik IM, Bialystok E. Planning and task management in older adults: Cooking breakfast. Memory & Cognition. 2006; 34(6):1236–1249. [PubMed: 17225505]

Crandall, B., Klein, GA., Hoffman, RR. Working minds: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press; 2006.

Crawford JO, Taylor C, Po NLW. A case study of on-screen prototypes and usability evaluation of electronic timers and food menu systems. International Journal of Human-Computer Interaction. 2001; 13(2):187–201. DOI: 10.1207/S15327590IJHC1302_6

Dattel AR, Durso FT, Bédard R. Procedural or conceptual training: Which is better for teaching novice pilots landings and traffic patterns? Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2009; 53(26):1964–1968. DOI: 10.1177/154193120905302618

Davis J, Rebelsky SA. Food-first computer science: Starting the first course right with PB&J. Proceedings of the 38th SIGCSE technical symposium on Computer science education. 2007:372–376.

Diaper, D. Understanding task analysis for human-computer interaction. In: Diaper, D., Stanton, N., editors. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004. p. 5-47.

Diaper, D., Stanton, NA. Wishing on a sTAr: The future of task analysis. In: Diaper, D., Stanton, N., editors. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004. p. 603-619.

Drysdale J, Casey J, Porter-Armstrong A. Effectiveness of training on community skills of children with intellectual disabilities. Scandinavian Journal of Occupational Therapy. 2008; 15:247–255. DOI: 10.1080/11038120802456136 [PubMed: 18855229]

Endsley, MR., Bolté, B., Jones, DG. Designing for situation awareness. An approach to user-centered design. Boca Raton, FL: CRC/Taylor& Francis; 2003.

Ericsson, KA., Smith, J. Prospects and limits of the empirical study of expertise: An introduction. In: Ericsson, K., Smith, J., editors. Toward a general theory of expertise: Prospects and limits. New York, NY: Cambridge University Press; 1991. p. 1-38.

Feldon DF, Timmerman BC, Stowe KA, Showman R. Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. Journal of Research in Science Teaching. 2010; 47(10):1165–1185.

Felipe, SK., Adams, AE., Rogers, WA., Fisk, AD. Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society; 2010. Training novices on Hierarchical Task Analysis.

Hickman JM, Rogers WA, Fisk AD. Training older adults to use new technology. Journals of Gerontology. 2007; 62B:77–84.

Hoffman, RR., Militello, LG. Perspectives on cognitive task analysis. New York, NY: Psychology Press; 2009.

Jonassen, DH., Hannum, WH., Tessmer, M. Handbook of task analysis procedures. Westport, CT: Praeger; 1989.

Jonassen DH, Tessmer M, Hannum WH. Task analysis methods for instructional design. 1999 Retrieved from netlibrary database.

Kieras, D. GOMS models for task analysis. In: Diaper, D., Stanton, NA., editors. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004. p. 83-116.

Kirwan, B., Ainsworth, LK. A guide to task analysis. London: Taylor & Francis; 1992.

Lane R, Stanton NA, Harrison D. Applying hierarchical task analysis to medication administration errors. Applied Ergonomics. 2006; 37:669–679. DOI: 10.1016/j.apergo.2005.08.001 [PubMed: 16182230]

Luker KR, Sullivan ME, Peyre SE, Sherman R, Grundwald T. The use of a cognitive task analysis-based multimedia program to teach surgical decision making in flexon tendon repair. The American Journal of Surgery. 2008; 195:11–15. [PubMed: 18082536]

Novick LR. Understanding spatial diagram structure: An analysis of hierarchies, matrices, and networks. The Quarterly Journal of Experimental Psychology. 2006; 59(10):1826–1856. [PubMed: 16945862]

Novick LR, Hurley SM. To matrix, network, or hierarchy: That is the question. Cognitive Psychology. 2001; 42(2):158–216. [PubMed: 11259107]

Olfman L, Mandviwalla M. Conceptual versus procedural software training for graphical user interfaces: A longitudinal field experiment. MIS Quarterly. 1994; 18(4):405–426.

Ormerod, TC., Shepherd, A. Using task analysis for information requirements specification: The sub-goal template (SGT) method. In: Diaper, D., Stanton, NA., editors. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004. p. 347-365.

Patrick J, Gregov A, Halliday P. Analysing and training task analysis. Instructional Science. 2000; 28(1):51–79.

Redish, JC., Wixon, D. Task analysis. In: Jacko, J., Sears, A., editors. The human-computer interaction handbook. Mahwah, NJ: Lawrence Erlbaum Associates; 2003. p. 922-940.

Schaafstal A, Schraagen JM, van Berlo M. Cognitive task analysis and innovation of training: The case of structured roubleshooting. Human Factors. 2000; 42(1):75–86. [PubMed: 10917147]

Scheiter K, Gerjets P, Schuh J. The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. Instructional Science. 2010; 38:487–502. DOI: 10.1007/s11251-009-9114-9

Seamster, TL., Redding, RE., Kaempf, GL. A skill-based cognitive task analysis framework. In: Schraagen, JM.Chipman, SF., Shalin, VL., editors. Cognitive task analysis. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2000. p. 135-146.

Shepherd A. An improved tabular format for task analysis. Journal of Occupational Psychology. 1976; 49:93–104.

Shepherd A. Hierarchical task analysis and training decisions. Programmed Learning and Educational Technology. 1985; 22:162–176.

Shepherd A. An approach to information requirements specification for process control tasks. Ergonomics. 1993; 36(11):1425–1437. DOI: 10.1080/00140139308968010

Shepherd A. HTA as a framework for task analysis. Ergonomics. 1998; 41(11):1537–1552. [PubMed: 9819574]

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

Shipley, W., Zachary, R. Shipley Institute of Living Scale. Los Angeles: Western Psychological Services; 1939.

Stanton NA. Hierarchical task analysis: Developments, applications, and extensions. Applied Ergonomics. 2006; 37:55–79. [PubMed: 16139236]

Stanton, NA., Young, MS. A guide to methodology in ergonomics Designing for human use. London, UK: Taylor & Francis; 1999.

Sullivan ME, Brown CVR, Peyre SE, Salim A, Martin M, Towfigh S, et al. The use of cognitive task analysis to improve the learning of percutaneous tracheostomy placement. The American Journal of Surgery. 2007; 193(1):96–99. [PubMed: 17188097]

van Merriënboer, JJG. Training complex cognitive skills. Englewood Cliffs, NJ: Educational Technology Publications; 1997.

Wechsler, D. Wechsler Memory Scale III. 3rd. San Antonio, TX: The Psychological Corporation; 1997.

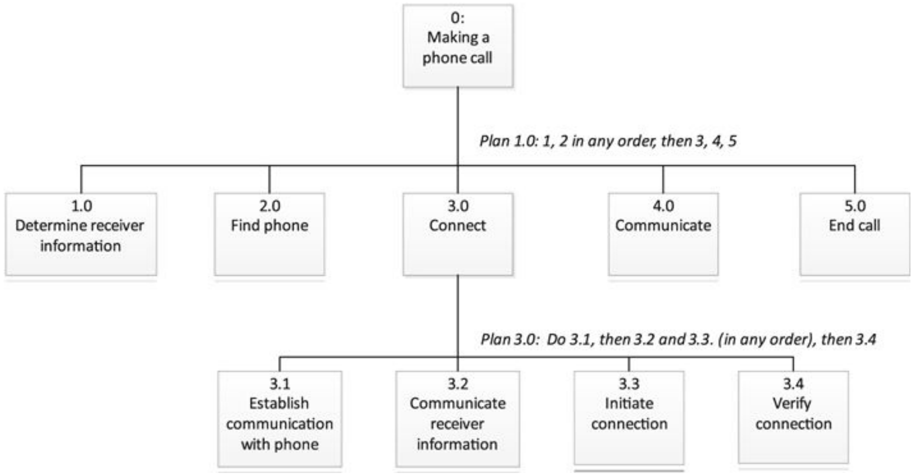

Fig. 1.

Example task redescription (breakdown of goals into subgoals, including plan). This example has a depth of two and a breadth of five

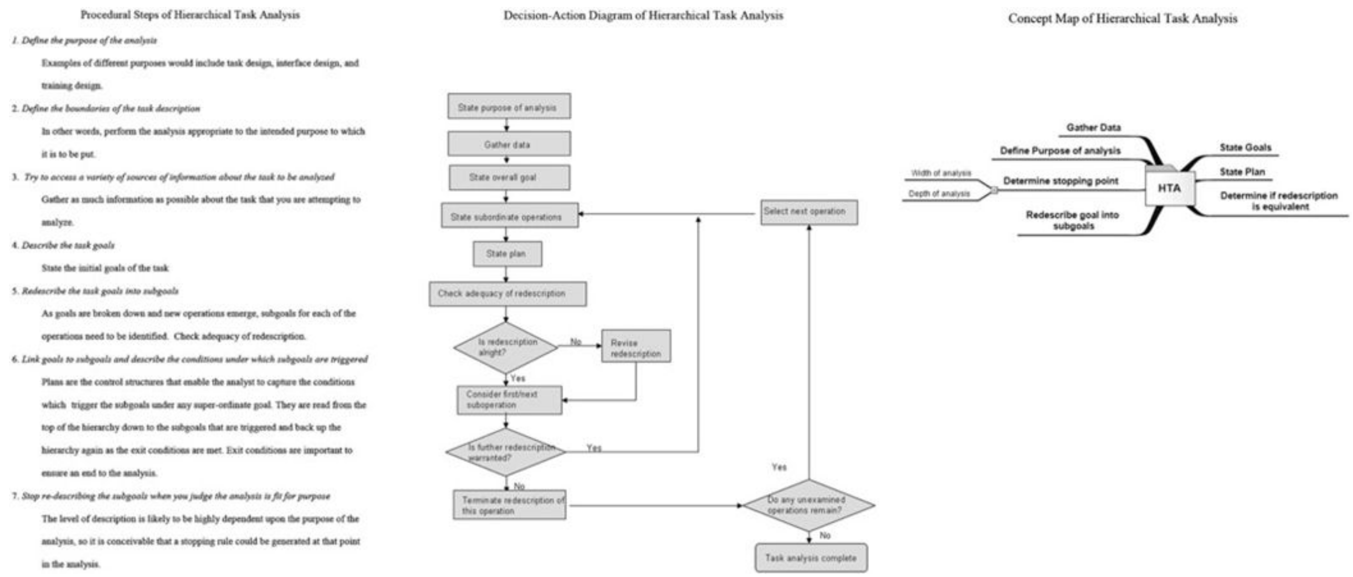

Fig. 2.

Additional instructions from Study 1

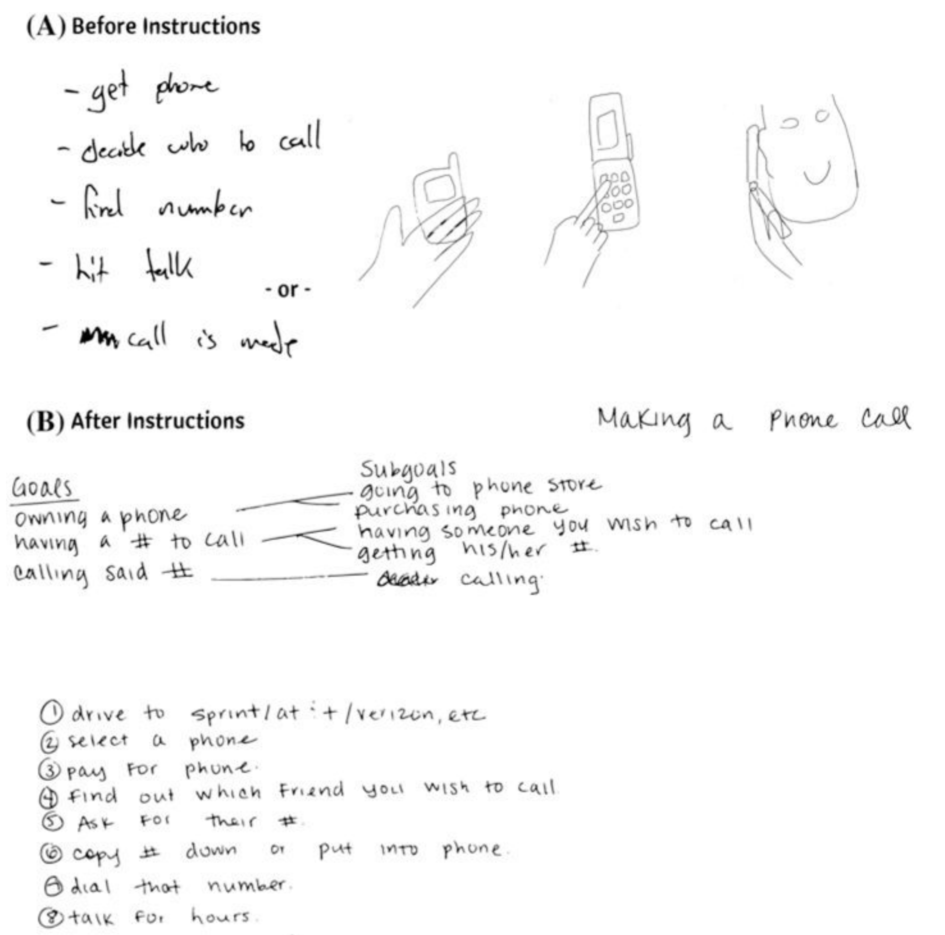

Fig. 3.

Example task analyses from Study 1 of novice participants before they read instructions (**a**) and of novice participants after they read the instructions (**b**)

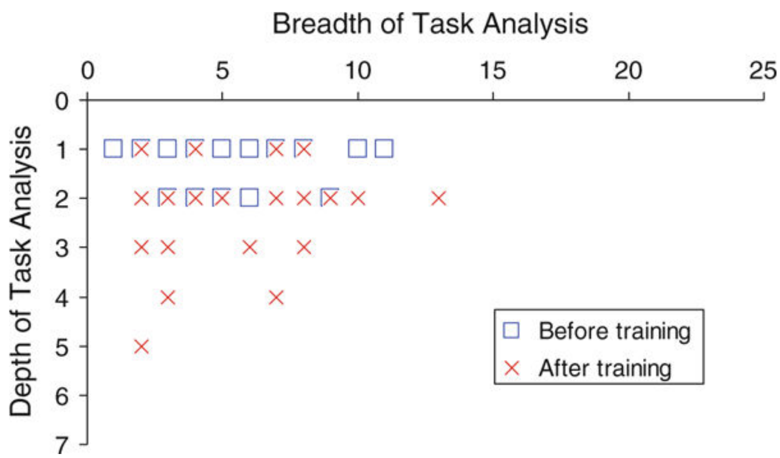

Fig. 4.
Study 1: Depth and breadth of task analyses for *making sandwich* and *making phone call* before and after training

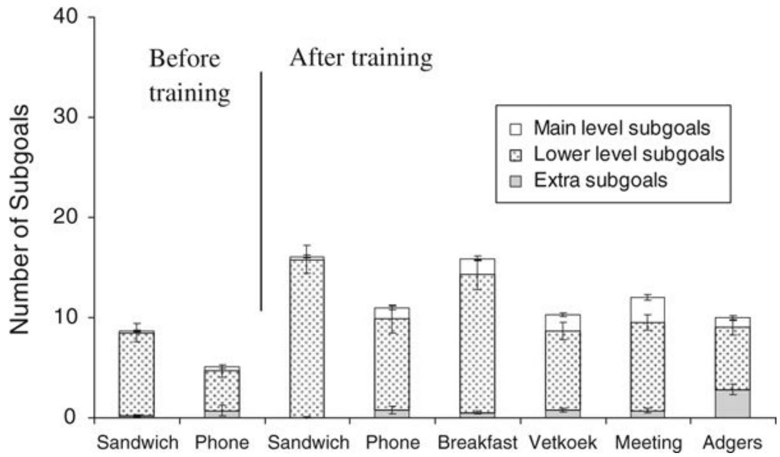

Fig. 5.

Study 1: average number and standard error for main level goals, lower level subgoals, and those not in the master task analysis (extra)

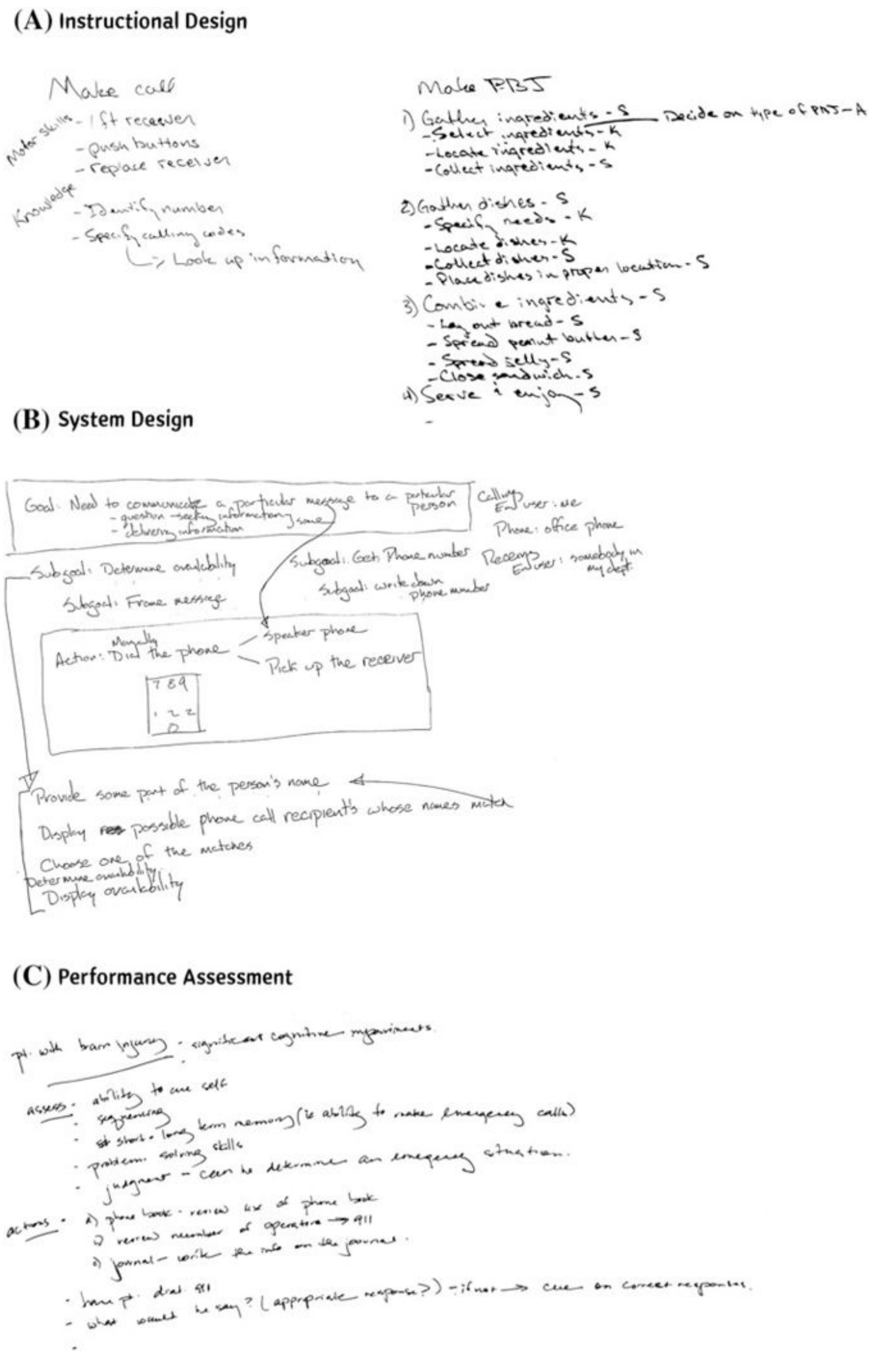

Fig. 6.
Example task analyses from Study 2 practitioners

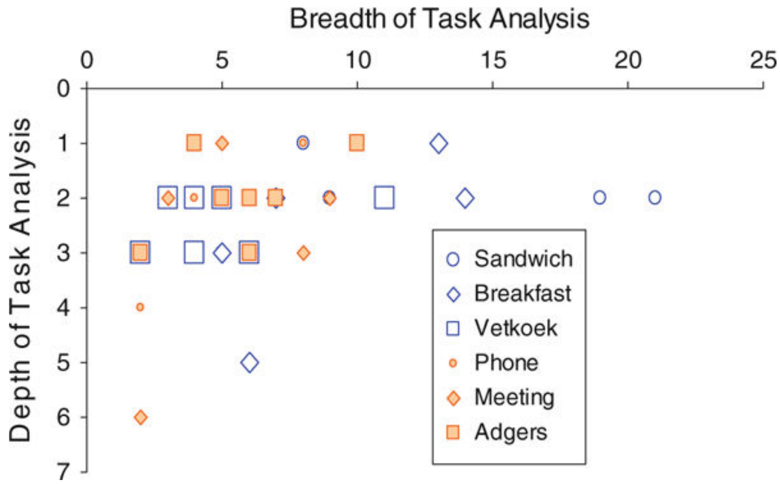

Fig. 7.
Breadth and depth of task analyses for all six tasks of Study 2 (*N* = 8 per task)

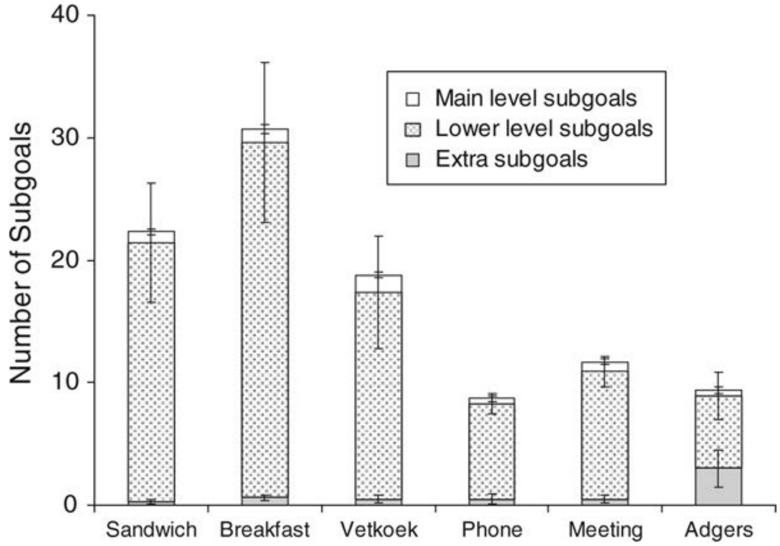

Fig. 8.

Study 2: average number and standard error for main level goals, lower level subgoals, and those not in the master task analysis (extra)

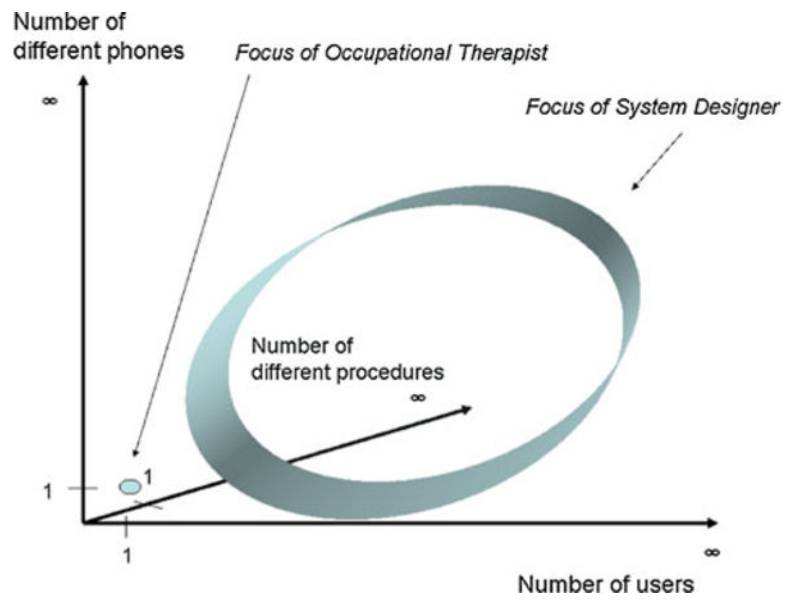

Fig. 9.

Possible task space of a system designer (*broad*) and an occupational therapist (*narrow*)

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Overview of the analyzed tasks

Domain	Familiarity	Specificity	Task as presented to participant
Cooking	Familiar	Specific	*Making* a peanut-butter jelly *sandwich*[a]
	Familiar	General	*Making breakfast*
	Unfamiliar	Specific	*Making Vetkoek* (a South African main dish)
Communication	Familiar	Specific	*Making* a *phone call*
	Familiar	General	*Arranging* a *meeting*
	Unfamiliar	Specific	*Sharing pictures* using Adgers[b] (a communication software)

[a]In the text, tasks are referred to by their shortened version as indicated in italics

[b]Adgers is a fictional product and no further details were provided

Table 2

Characteristics of Participants in Study 1 (N = 12 per condition)

Measure	**Training condition**						***p***
	Steps		**Decision-action**		**Concept map**		
	M	***SD***	***M***	***SD***	***M***	***SD***	
Age	20.42	1.78	20.50	1.62	20.83	1.19	–
Digit symbol substitution[a]	95.33	12.71	95.33	15.59	105.08	15.87	.19
Reverse digit span[b]	8.92	1.88	8.58	2.02	7.75	1.66	.30
Shipley vocabulary[c]	32.75	3.14	32.67	3.17	32.00	3.46	.83

Alpha level was set at .05; none of the group differences were significant

[a]Number correct out of 120 within 2 min (Wechsler 1997)

[b]Number correct out of 14 (Wechsler)

[c]Number correct out of 40 (Shipley and Zachary 1939)

Table 3

Study 1 coding scheme for task analyses and questionnaires

Feature of HTA	Coding procedural knowledge	Coding declarative knowledge
Format/Plan	(a) In what format was sequence expressed? (e.g., words, numbers, flow chart) (b) Was the label *plan* used?	State plan (sequence of events) to show when to carry out subgoals, e.g., “Have a plan”
Hierarchy	(a) What was the breadth of the task analysis on the highest level? (b) What was the depth of the task analysis at its deepest?	It is a hierarchy of goals and subgoals
Main goal	Was the main goal stated?	State the high-level goal, overall goal to be achieved
Subgoals	(a) What subgoals were identified? (b) Was the label *subgoal* used?	The elements necessary to carry out the high-level goal, e.g., “State subgoals”
Criteria	Were criteria mentioned to help determine whether the goal was reached satisfactorily?	The satisfaction criteria that help establish if the task has been properly completed, e.g., “Ensure that final goal is satisfied”
Specificity	Was the task analysis general or specific, that is, did it account for three or more task variations?	Not included in instructions
Other	Not applicable	Does not fit any other category

Table 4

Breadth and depth of task analyses

	Breadth				Depth			
	M	SD	Min	Max	*M*	SD	Min	Max
Study 1: Novices								
Cooking								
Sandwich before training	5.50	2.36	3	11	1.33	.49	1	2
Sandwich after training	5.94	2.94	2	13	2.17	.79	1	4
Breakfast	4.53	3.57	1	15	2.14	.83	1	4
Vetkoek	3.67	1.53	1	7	1.92	.84	1	4
Communication								
Phone before training	4.22	1.70	1	8	1.11	.32	1	2
Phone after training	4.06	2.44	2	10	2.33	1.19	1	5
Meeting	4.25	2.21	1	12	2.06	.79	1	4
Adgers	4.44	2.08	2	10	2.14	.90	1	4
Study 2: Practitioners								
Cooking								
Sandwich	10.13	6.60	4	21	2.00	.53	1	3
Breakfast	6.75	4.46	2	14	2.50	1.20	1	5
Vetkoek	4.88	2.75	2	11	2.38	.52	2	3
Communication								
Phone	3.38	2.00	2	8	2.38	.92	1	4
Meeting	6.25	2.87	2	9	2.38	1.51	1	6
Adgers	5.25	2.66	2	10	2.13	.83	1	3

Table 5

Overview of findings from Study 1: training implications

HTA feature	Desirable outcomes	Areas of concern	Implications for training
Hierarchy	Instructions helped increase task analysis depth. Average breadth on highest level of analysis was within 3–8 subgoals	Hierarchical was not seen as a main feature of HTA, and some participants kept a depth of one Some task analyses were too broad or too narrow	Emphasize task analysis depth Discuss implications of task analysis that is too narrow or broad
Main goal	Novices stated the main goal	Novices adjusted the main goal	Discuss implications of adjusting the main goal
Subgoals	Instructions were helpful for emphasizing the concept of subgoal. Novices identified more subgoals after training	Novices kept their focus on lower level subgoals	Show novices how to derive higher-level subgoals from lower level ones
Format/Plan	Instructions were helpful for concept of plan	Novices mix and match formats and only define one plan.	Start with list-style format, then move to hierarchy
Criteria	Novices included criteria after training	Few novices listed satisfaction criteria as a feature of HTA.	Emphasize the importance of satisfaction criteria
Versatility	HTAs increased in versatility across tasks	Task analyses remain specific	Emphasize and illustrate how a versatile (general) task analysis differs from a specific one

Table 6

Study 2 participant characteristics ($N = 8$)

	M	SD	Min	Max
Age	39	8.6	27	54
No. of TA in past year	12.8	17.7	2	50
No. of TA in lifetime			<5	>50
No. of methods reported[a]	4	3	0	8

[a]Data from seven participants only

Table 7

Skill components and guidelines for evaluating task analysis

Objective	Skill components	Guideline
Determine task goal	Determine purpose of the task analysis Set scope of analysis State assumptions Differentiate purpose and goal Delineate task to be analyzed (main goal) from the task of conducting a task analysis Identify goal Define goal of the task Question the given goal Determine super-ordinate goal Delineate task goal from other (similar) task goals Ask question	Spend time discussing activities at the onset of the task analysis: main goal (task) and the purpose of the task analysis
Determine plan	Separate sequence from content Determine sequence Choose format Follow format Adjust format	Prepare for a variety of task analysis formats
Create hierarchy	Expand breadth Reduce breadth Expand depth Reduce depth Adjust subgoal placement Decide whether to include or exclude a subgoal Determine location of subgoal Set task structure dimensions Evaluate subgoal placement Move element within same level Place subgoal in hierarchy Move element to another level	Emphasize that developing a hierarchy is an active process that involves creating and expanding a task space, and negotiating the placement of a subgoal
Determine task boundaries	State assumptions Define breadth of task analysis Determine breadth of task analysis Determine if breadth is appropriate Recall minimum breadth Recall maximum breadth Define depth of task analysis Determine depth of task analysis	Be prepared to provide definitions, as depth and breadth of task analyses may be influenced by definitions or a process
State subgoals	Ask questions Consider prior knowledge Consider task factors Consider time constraints Check symmetrical subgoals Check task aspects Refer to task constraints Search task space Specify requirements State assumptions Understand task space Determine procedure Identify objects Select level of analysis Define subgoal Identify subgoal Delineate subgoal from other subgoals Determine exact wording Notice if subgoal is outside boundaries	Employ a variety of methods to redescribe subgoals, and beware that factors other than the task to be analyzed may be important
Create versatility	Consider different ways to complete a task Consider task variations Constrain task space Create scenario Determine task variables Refer to task analysis purpose Refer to task performance Select task variables	Purposefully consider task variations to make a task analysis more general, and state your assumptions when you choose specific instances

Objective	Skill components	Guideline
Determine criteria	Identify criteria State criteria	Emphasize satisfaction criteria so they are not overlooked