An Evaluation of Tailored Web Materials for Public Administration

Nathalie Colineau & Cécile Paris
CSIRO – ICT Centre
cnr. Vimiera and Pembroke Roads
Marsfield, NSW 2122, Australia
(+61) 2 9372 4222
{nathalie.colineau,cecile.paris}@csiro.au

Keith Vander Linden
Department of Computer Science
Calvin College
Grand Rapids, MI, 49546, USA
(+1) 616 526 7111
kvlinden@calvin.edu

ABSTRACT
Public Administration organizations generally write their citizen-focused, informational materials for generic audiences because they don’t have the resources to produce personalized materials for everyone. The goal of this project is to replace these generic materials, which must include careful discussions of the conditions distinguishing the various constituencies within the generic audience, with tailored materials, which can be automatically personalized to focus on the information relevant to an individual reader. Two key questions must be addressed. First, are the automatically produced, tailored forms more effective than the generic forms they replace, and second, is the time the reader spends specifying the demographic information on which the tailoring is based too costly to be worth the effort. This paper describes an adaptive hypermedia application that produces tailored materials for students exploring government educational entitlement programs, and focuses in particular on the effectiveness of the generated tailored material.

Categories and Subject Descriptors
H.5 [Information Interfaces and Presentation]: General; I.2.7; [Natural Language Processing]: Language Generation.

General Terms
Documentation, Human Factors.

Keywords
Tailored Information Delivery, Natural Language Generation, Public Administration.

1. INTRODUCTION
Citizen-focused materials in Public Administration (PA) must describe public programs and present the conditions under which individuals or organizations are eligible for them. Because PA constituencies are diverse, these conditions and associated informational materials are notoriously complicated. While PA organizations could assign a personal case worker to each citizen who would, like a personal physician, communicate directly with that citizen, cost and resource constraints generally force the organizations to communicate with their constituency en masse using generic materials. Unfortunately, generic materials tend to be longer and harder to understand than materials tailored for particular readers.

This paper describes an application that automatically produces tailored informational materials. The application represents and reasons about conditions and concepts using semantic web technologies and generates tailored materials using discourse planning technologies. This application has been developed in collaboration with Centrelink, the Australian service-delivery organization for Human Services, and focuses on materials for students exploring educational entitlement programs.

As an example of the sort of PA materials addressed in this work, consider the website thatCentrelink publishes for its Youth Allowance (YA) program, an entitlement program that helps Australian students with the cost of studying [4]. This website is targeted at a generic audience. Figure 1(a) shows an excerpt that details the eligibility conditions for the program (i.e., “You may be eligible for Youth Allowance if…”). In order to cover all possible student readers, this webpage includes a rather complex set of conditions with overlapping age ranges (i.e., 18-24, 16-24, 16-20, 16-17 and 15) and carefully-worded logical explanatory texts corresponding to each condition. Compare this with the excerpt from the tailored website generated by the prototype application described in this paper; see Figure 1(b). Here, the tailored text covers the same basic conditions but has been tailored for a reader who has indicated that she is a 19-year-old Australian resident planning to attend university full time. Given the context of this reader, the prototype has replaced the rather complex eligibility text with a one-paragraph summary of the eligibility conditions the reader is known to satisfy (“You’ve told us that you meet the following criteria: …”). A similar simplification occurs later in the document in the materials related to the reader’s likely payment rates. While the generic site for Youth Allowance must list all eleven possible payment rates, the tailored site needs only to list the one rate relevant for the given reader.

There are two key questions with respect to tailoring applications such as the one described here. First, will the tailored version of the material be sufficiently clear to enable readers to comprehend the content more accurately than they currently do with the generic versions? Second, will the extra time that the reader must spend specifying the demographic information on which the tailoring is based slow them down in comparison to the generic website, which requires no such information? This paper addresses these questions by reviewing relevant work, describing an application that produces tailored materials with a particular emphasis on the forms of tailoring it produces, and presenting the results of a pair of studies designed to evaluate the relative
comprehensibility and efficiency of the generic and the tailored websites.

2. RELATED WORK

The Adaptive Hypermedia (AH) prototype application described in this paper is based on technologies from Document Automation (DA), Natural Language Generation (NLG) and Ontology Verbalization (OV). The types of tailoring it generates are inspired by tailoring research in eHealth.

2.1 Tailoring in DA and NLG

This work draws on the practice of current DA systems (e.g., HotDocs, Exari, and Arbortext)\(^1\) and on the theory and techniques developed in NLG [14] [13].

DA tools have been used in the legal profession to automate the production of custom-built legal documents (e.g., deeds of sale, standardized agreements, etc.) and in the technical documentation field to produce model-specific product documentation. These tools provide mail-merge-like features extended with conditional inclusion/exclusion of coarse-grained text units generally on the order of sections, paragraphs or perhaps sentences. The work reported in this paper can be seen as part of the broader field of DA but, as will be shown, deploys more sophisticated ontological representations [6] and planning and generation techniques [7].

Tailoring systems have been studied extensively in NLG. For example, the STOP system developed by Reiter et al. (2003) produced tailored smoking cessation letters using a combination of tailoring techniques, though Reiter and his colleagues were unable to demonstrate the effectiveness of the tailored letters in a large field study within their target audience [15]. DiMarco and her colleagues developed HealthDoc, which generates tailored health-education documents from a master document [8]. In the domain of corporate communication, the SciFly system developed by Colineau and Paris produced tailored informational brochures and succeeded in demonstrating effectiveness in its evaluation [5]. The work presented in this paper manages more coarse-grained text units than do STOP and HealthDoc and thus addresses fewer issues in sentence-level generation. The work follows more directly from SciFly in that it is based on the same implementation platform [13], a platform whose architecture is built around a hierarchical text planner [11].

2.2 Adaptive Hypermedia (AH)

A variety of techniques, including those from DA and NLG, have been applied to the adaptive presentation of hypermedia content. Using the terms and categories defined by Kobsa et al. (2001) in their survey of AH [10], the application described here collects user-supplied demographic data using an on-line questionnaire, which is commonly done in fielded systems, and then tailors the bulk of its content using a fragment-variant approach, which selects and assembles pre-authored fragments of text. This approach works well for the carefully-worded, legal content commonly found in citizen-focused PA texts, where generating

---


---

Figure 1. (a) an excerpt from the current website; (b) a corresponding excerpt from the tailored website.
sentences on-the-fly from first principles is not possible given that all material has to be legally approved before release.

The application represents and reasons about its users and its content using a knowledge base implemented with semantic web technologies [16], which can be seen as what Bunt et al. would categorize as domain-dependent, abstract information [2]. Of particular concern in citizen-focused PA texts is the notion of eligibility. The application uses OWL-DL [12] to represent and reason about eligibility conditions in order to determine both whether a user is eligible for a particular program and what content to include.

2.3 Ontology Verbalization (OV)
The application’s use of OWL-based representations of eligibility conditions makes it possible to use OV techniques, which take logical expressions written in an ontology language and generate corresponding textual descriptions [17]. This approach tends to implement structural mappings from all possible ontological expressions to simple textual forms that are accurate but perhaps not as fluent as might be desired.

The work described here implements OV for the subset of OWL expressions most commonly used in citizen-focused PA documents. To improve the fluency of the output, it also provides a mechanism for authors to override the automated expression when logical expressions become too complex to produce sufficiently fluent texts or when the texts must be more carefully crafted for legal purposes. This tool, which is not addressed in this paper, is a frame-based authoring tool (cf. [1]) that allows authors to specify domain content, including conditions and text. More details on this tool can be found elsewhere [7].

2.4 Tailoring in E-Health
While there is surprisingly little work on tailored communication in PA applications, work on tailoring has become a common theme in eHealth applications [3]. In a recent meta-study of tailored health documents [9], Hawkins et al. report that tailoring in eHealth has generally proven to be effective, particularly when there is a relatively large variation in the readers. The PA domain serves a similarly diverse constituency that varies on a number of scales, including age, race, domicile, family status, etc.

Hawkins et al. identified the following three forms of tailoring as being commonly effective in eHealth applications:

- **Personalization** – A document should identify itself to its reader as a tailored document, for example, by saying things like “this document has been produced for you…”
- **Feedback** – A document should refer to information provided by the reader, for example, by saying things like “You’ve told us that…”.
- **Content matching** – A document should include the most appropriate content for a particular reader, for example, by choosing information units based on the reader’s personal information. This comprises a variety of content selection techniques.

The prototype adaptive hypermedia application described here deploys all three forms of tailoring.

Given that these forms were developed in eHealth rather than PA, it was not clear that they would be equally effective in PA. One difference between typical eHealth applications and the PA application discussed in this paper is that eHealth readers must be convinced to change their lifestyle, say to stop smoking or eat more vegetables, which can be difficult. In contrast, PA readers are looking for government financial assistance and are highly motivated to sort out eligibility conditions.

3. APPLICATION
The prototype application, whose architecture is shown in Figure 2, comprises a web-based query wizard, which collects demographic information about the reader, the Document Planner, which uses this information to select and structure web materials that are tailored for the reader, and a Content Model, which represents and reasons over the information content.

3.1 Query Wizard
The query wizard, shown on the left of Figure 2, uses a simple jQuery-based questionnaire to collect demographic information from the reader, including their age, citizenship, race, school plans, and living arrangements, see Figure 3. It includes all the user information required to make eligibility and tailoring decisions throughout the website.

The wizard stores the user information in the Content Model and then asks the Document Planner to select and present a suitably tailored presentation of the material relevant to the context. If the reader provides no information, then the planner returns a full,
The more information the reader provides, the more feedback and content matching the planner delivers.

In the early phases of development, it was unclear how much information readers would be willing to provide through the query wizard. So a preliminary study, described below, evaluated the relative desirability of a short query wizard, with a minimal set of four questions, and a longer query wizard, with more questions but resulting in more highly tailored material. As detailed below, readers preferred the longer wizard. The question then became whether or not the cost of filling out the longer query wizard would be outweighed by the value of the resulting tailored material. This was the focus of the second study described here.

The application has the ability to automatically plan wizard questions using OV techniques, but the system tested in this study used the hand-authored, JQuery wizard shown in Figure 3. Details on the OV query mechanism can be found elsewhere [7].

3.2 Document Planner

The Document Planner, shown in the upper-right of Figure 2, plans the document structure and content using a text planning engine driven by declarative plan operators. The plan operators are designed to build hierarchical text structures with conditionally included constituents. This technology is well-known in NLG [14], and its use in the Myriad platform, the platform underlying the application, is documented elsewhere [13].

The prototype exhibits one key departure from traditional approaches to building NLG applications in that it is a hybrid system that works with both coarse-grained information units on the order of sections, paragraphs and sentences, termed fragment variants in the AH community [10], and the fine-grained information units on the order of words as is more common in NLG systems. Traditional NLG applications include tactical generation components that work exclusively with units at the lexical and sentence levels. This application structures the output by combining hand-written text fragments where available and using finer-grained OV where necessary.

This hybrid approach to NLG implements the three forms of tailoring discussed above as follows.

Personalization is simple enough that it requires only basic information on how to address the reader. For example, the tailored output for the YA program webpage opens with the following statement:

“This web material has been produced for you on 19/Jan/2012 10:29AM based on the information you have provided.”

This basic form of tailoring has proven to be effective as the material is more likely to be read, remembered, and considered as interesting if people believe it has been produced for them. The planner uses a simple template-filling mechanism to produce this text, and places it on the webpage that links to the one shown in Figure 1(b).

Feedback is more involved, requiring finer-grained content and more detailed generation techniques. For example, the portion of the Youth Allowance brochure tailored for a reader known to be a 21-year-old Australian citizen working as an apprentice includes the text shown in Figure 1(b). The planner uses simple templates to produce the opening and closing texts in this example (i.e., “You’ve told us that…” and “so you should be eligible…” and OV to produce the declarative conditional expressions listed in the bullets. It uses the reasoning capabilities of the content model (see the next section) to determine the eligibility conditions the reader is known to satisfy, in this case age, residency and study conditions, and then constructs the feedback paragraph for the reader as appropriate. Had the reader’s circumstances on any of these conditions been unknown, the planner would have included full descriptions of the unknown conditions in the tailored output. For example, Figure 4 shows the eligibility page tailored for a reader who is known to be an 18 year-old Australian citizen but whose course of study is unknown. The age and citizenship

![Figure 3. The query wizard](image-url)
conditions are mentioned only briefly, while the full explanation of the course requirements is given. The assumption is that if readers do not specify a condition, they will need a full description of that condition.

**Content matching** in the prototype application is based on eligibility. The Content Model can represent and reason about eligibility conditions on any content element at any level of abstraction, which allows the document planner to include only those content elements that are relevant to the current reader. For example, the 21-year-old, Australian university student addressed by the text in Figure 1(b) is not eligible for the ABSTUDY program because she is not an indigenous Australian nor for the Austudy program, because she is too young. So the application does not include any discussion of either of these programs. It focuses instead on the one program for which the reader is eligible, Youth Allowance. The hope was that this would help readers determine their eligibility more accurately.

Similarly, the planner can match the appropriate payment rate to a given reader, a computation that requires choosing among potentially many possible rates based on the reader’s circumstances. In the case of Youth Allowance, there are eight rates from which to choose. With the information about the reader collected by the Query Wizard, the prototype can select one rate to present. The hope was that this would help readers determine their potential payments more accurately.

After the reader completes the wizard, the planner produces a list of student programs for which the reader may potentially be eligible based on the information they have provided. If the reader provides no information, then the list includes all the programs, as shown in Figure 5. Here, all three programs are listed: Youth Allowance (YA), Austudy and ABSTUDY, each with a short hand-authored description fragment and a list of conditions for which the reader’s information is unknown (i.e., “Note: The following criteria, for which we don’t know…”). The Content Model includes both short and long versions of the program descriptions which the Document Planner deploys as appropriate (cf. Kobsa et al.’s notion of fragment variants [10]). The Content Model also includes OWL-DL representations of each of the listed conditions and an optional hand-authored text for the description. The Document Planner chooses the hand-authored version when available and otherwise reverts to OV for the condition. For example, Figure 4 shows two conditional expressions produced with OV (i.e., “You are under 25…” and “You are an Australian citizen…”) and one hand-authored condition expression (i.e., “You must be undertaking an approved course or Australian Apprenticeship.”). The OV expressions are relatively simple to produce automatically. Conjunctions and disjunctions, however, such as the one shown in this hand-authored text, are a well-known problem area for NLG systems, and while the prototype application can use OV to generate the course condition (i.e., “You must be undertaking an approved course and you must be…”), the hand-authored version is easier to read. This is the motivation for adopting a hybrid solution to NLG in the prototype application.

Note also that the OV mechanism is capable of producing both declarative forms (e.g., “you are under 25…”) and obligatory forms (i.e., “you must be under 25…”). The discourse planner chooses the appropriate form based on whether the reader’s circumstances are known or unknown.

### 3.3 Content Model

The Content Model, shown on the lower-right of Figure 2, represents the conceptual structure of the domain, the user model and the text fragments for all the entities being expressed. It is implemented using OWL-DL [16] and supports the reasoning required for tailoring discussed above using Hermit [12].

To support reuse of the various elements of this model by other administrative organizations or by other tailored delivery applications, the model is divided into the three separate but integrated ontologies:
The Myriad ontology represents those entities and relationships required for the document planner, e.g., the Person entity and the hasAge relationship.

The Government core ontology represents general PA elements, e.g., Program and hasCondition.

The Centrelink ontology represents those elements specific to the particular organization being models, e.g., Centrelink’s Youth Allowance program.

More details on this structure can be found elsewhere [6].

These three ontologies work together to support the full context of communication. This includes: (1) a user model, which represents the demographic information provided by the reader through the Query Wizard; and (2) a domain model, which represents the content fragments, domain entities and eligibility constraints. The user and domain models are not explicitly shown in Figure 2. The Content Model also supports the representation of a device context, the discourse history and the task context, but these are not deployed in this application [13]. This structure is designed to minimize the effort required to port the knowledge base to new domains, but the true cost of this in the general case is unclear; a preliminary study of this issue is reported elsewhere [13].

The prototype supports the eligibility reasoning required for tailoring decisions using categorical reasoning as implemented by HermiT [12], a reasoner for OWL ontologies. For each reader, the prototype asserts an individual of class Person and sets the properties for that individual as appropriate, e.g., race, age, etc. It then uses HermiT’s classification to determine if the individual’s properties satisfy the properties required for any of the programs or other content elements. The results of this reasoning are provided to the Document Planner for use as described above.

The text shown in Figure 5 also includes hyperlinks to definitions of the more complex conditions (i.e., through the hypertext link on “(details)”). For example, the definition of indigenous Australian includes the following text:

An Indigenous Australian is someone who:

- is of Australian Aboriginal or Torres Strait Islander descent
- identifies as an Australian Aboriginal or Torres Strait Islander person, and
- is accepted as such by the community in which they live, or have lived.

You may be asked to provide evidence to prove you meet all three parts of this definition.

This important legal text is socially and politically sensitive and must, therefore, be carefully worded. This is the primary motivation for including hand-authored text fragments in the Content Model and for using hybrid technologies for text planning and generation.

Details on the nature of this model and the reasoning it supports, particularly with respect to PA and information sharing, can be found elsewhere [6] [7].

4. EVALUATION

Two evaluations of this application have been conducted: first, a qualitative evaluation and second a more quantitative study. This section details both.
4.1 A Qualitative Study of User Impressions

Centrelink’s Concept Lab, an independent evaluation team within Centrelink, ran a qualitative evaluation of the first version of this application in May 2011. The goal was to evaluate the desirability of the tailored forms and their presentation in a brochure format, and to assess whether citizens would be willing to provide the personal information required to receive tailored information. There was particular interest in ascertaining how detailed a questionnaire the readers would be willing to answer, so both a short and a long form were provided. The short form included only four questions; the long form included twelve. These original forms differed slightly from the current form of the questionnaire shown in Figure 3. The test website also provided both web-formatted and paper-formatted output, all presented electronically.

4.1.1 Method

The evaluation included 15 participants, all of whom were of an age appropriate for Youth Allowance (YA) and of whom 14 were currently in receipt of YA payments. The materials included information on the YA, ABSTUDY and Austudy programs.

During 30-minute sessions, the participants explored the prototype, which included both the short and long forms of the query wizard and produced the program list as well as both the web and paper versions of the program materials. The free flowing nature of the subjects’ exploration was not conducive to asking a structured set of questions. Instead, all participants were encouraged throughout the sessions to provide their preferences and feedback as to which wizard they preferred and if they were in favor of the concept of tailoring. The feedback was consistent throughout all 15 sessions.

4.1.2 Results

During the sessions the participants explored the prototype and tested the two different pathways: the “quick” wizard and the “detailed” wizard. Those who chose first the quick wizard thought it would be the fastest way to collect relevant information but later realized that the detailed wizard, through a more comprehensive question set, was providing them with more relevant content. All participants were able to answer the questions of the query wizard with ease, and all indicated that their structure was clear, concise and easy to follow.

All the participants expressed a preference for the tailored version of the brochure over the automatically-generated untailored version, with a preference for the longer more detailed form of the wizard. They even asked for a greater number of questions provided that it would lead to a greater level of tailoring.

Some suggested that we rephrase some of the questions, e.g., ‘are you intending on studying’, so that they could also use the tool to run different scenarios and understand under which conditions they could potentially qualify for Centrelink student payments.

The participants also expressed preferences for each of the three forms of tailoring deployed in the materials when they noticed them. All participants who noticed and commented on the personalization (e.g., “The following content sections have each been produced for you.”) liked it. All participants noticed the feedback in the tailored eligibility section and immediately indicated that they liked it. Finally, all participants preferred the shorter materials made possible by tailored content elision (e.g., the elision of the full definition of race in the eligibility section).

The results of this study generally supported the desirability of tailoring in this domain, its value, the acceptability of the query wizard approach and provided some helpful feedback on expression forms and structure.
4.2 A Quantitative Performance Comparison

Given the favorable reception of the application’s brochure output in the first evaluation, the prototype application was upgraded based on the subjects’ comments, and a second evaluation was conducted. The objective of this second study was to evaluate whether providing tailored PA web materials helps readers find the information they need more accurately and efficiently. Where the first study included both web and paper-brochure-formatted output forms, the second study focused on web output and conducted a direct comparison between the prototype’s automatically tailored website and Centrelink’s existing website. The study was conducted from November 29th through December 8th, 2011 at Calvin College (USA).

4.2.1 Method

We compared two sets of webpages: 1) a set of generic pages corresponding to an extract of the current Centrelink website about student payment programs, and 2) a tailored version of the generic pages. The first page of each of these websites is shown in Figure 6. With the current generic version, see Figure 6(a), readers have to choose a particular student program, then sift through the eligibility conditions to determine whether they are eligible or not for this program, and then the payment conditions to determine the amount they may get if any. With the tailored version, see Figure 6(b), readers have to answer a short questionnaire, as described above in Section 3.1, to inform the system about their situation, and, based on this information, the system tells them the program(s) they may be eligible for and the amount they may be paid under each program. The format of both test websites is consistent and based on the current government website with extraneous links and information removed for the purpose of testing.

We tested these two online versions with a group of 28 students aged 18-20 from the introductory programming course at Calvin College. These students were very much like the recipients of Youth Allowance demographically, except that they were US and Canadian citizens rather than Australian residents (e.g., similar age, currently going to school full-time at an accredited university or college, single with no children). They received course credit for being part of the study and knew nothing of Myriad (the underlying system tested), Centrelink (the Australian government agency providing these student programs), Youth Allowance or the project in advance. The students were split into two groups of 14, so that we could test each student for both versions and account for order effects. To counterbalance the effect of the order in which the systems were tested, we had group 1 evaluate first the tailored version and then the generic version, while group 2 started first with the generic version and then moved to the tailored version.

When testing each version, we asked students to determine: (1) whether they were eligible for any of the student payment programs; and (2) how much payment they may receive. We measured the accuracy of their answers and recorded the time it took them to answer each question. Then, we asked students to rate how confident they were about their answers, using a 6 point-scale from 0 (completely unconfident) to 5 (completely confident). To analyze the results, we conducted an analysis of variance using a mixed model with two fixed effects (i.e., group and tailoring) and one random effect of subjects. In all the results described below, group 1 (G1) refers to the group that saw the tailored version first, and group 2 (G2) to the group that saw the existing generic website first.

4.2.2 Results

Determining eligibility – The first question asked to students was to determine whether they were eligible for any of the student programs. We recorded their answers encoding them as 1 for correct and 0 for incorrect. As shown in Figure 7, both groups performed better with the tailored material. We noted that the difference in performance is more noticeable for group 1 who made almost twice as many mistakes with the non-tailored version. The results of the ANOVA confirmed that the difference between the generic and the tailored versions was statistically significant ($F(1, 26) = 6.56; p < 0.025$).

![Figure 7. Plot of the correctness of answers’ mean against the tailoring effect](image)

We recorded the time (in seconds) the students needed to determine their eligibility. The difference in task performance varies widely from one student to another ranging from 12 to 268 seconds. We also observed that the task performance varies a lot in the generic version from the first trial to the second with a group mean ranging from 106 to 51 seconds. Comparing the performance between the two trials (see Figure 8), we noticed that regardless of the version tested, students seemed to be faster when performing the task a second time (i.e., in the second trial).

![Figure 8. Plot of the time mean against the trial effect](image)
We conducted an ANOVA which confirmed that there was a statistically significant effect of order ($F(1, 26) = 5.93; p < 0.025$). The familiarity with the task may have played a role here, although as seen previously, this did not have any effect on the accuracy of their responses.

Determining the payment rate – The second question asked to students was to determine the fortnightly rate they may receive. (Note, being eligible to a student payment program did not guarantee students would be paid anything, as this depends on their income or that of their family). As shown in Figure 9, we observed here a larger difference in scores between the tailored and the non-tailored version with students making on average twice as many mistakes with the non-tailored version. The ANOVA results confirmed that the difference observed was statistically significant ($F(1, 26) = 7.55; p < 0.02$).

![Figure 9. Plot of the correctness of answers’ mean against the tailoring effect](image)

We recorded the time students needed to answer the second question. As previously observed, we noted differences in task performance between students, ranging from 9 to 266 seconds, with an overall average of 79.5 seconds. As shown in Figure 10, students tended to be faster when performing the task the second time regardless of the version tested. This difference is statistically significant ($F(1, 26) = 4.74; p < 0.05$), demonstrating an effect of order.

![Figure 10. Plot of the time mean against the trial effect](image)

Level of Confidence – Finally, comparing how confident students were in relation to their answers, we did not find any effect of tailoring ($F(1, 26) = 1.95; p > 0.05$). Although G2 seemed to be more confident with the tailored version, participants in G1 were almost as confident with both versions. We noticed however that both groups seemed to be more confident when performing the task the second time, as shown in Figure 11. This difference, although quite large for G2, was not significant ($F(1, 26) = 3.83; p > 0.05$). Therefore, regardless of the time spent and mistakes made, students were reasonably confident about their answers.

![Figure 11. Plot of the confidence level mean against the time effect](image)

Preferences – At the end of the test the subjects were asked reflective questions on which website they preferred and which made their job easier (“Did you prefer one version of the system over the other?”; “Which system made the tasks easier to do?” and “How useful was the questionnaire (wizard) in guiding you through the process?”). The participants overwhelmingly preferred the tailored website, and found that it made the task easier (see Figure 12). When asked about the usefulness of the tailored website’s wizard, using a 4 point-scale from 0 (not useful) to 3 (very useful), the average response was useful (2.0).

![Figure 12. Answers to the reflective questions](image)

4.2.3 Discussion

This study shows that tailoring the information presented to students helped them determine with more accuracy whether they were eligible to a particular student assistance program and how much they may receive. The tailored version of the system did not help student find the information more efficiently, but considering that students had to complete the Query Wizard for the information to be tailored to their circumstances, it is sufficient to know that it did not penalize them either. The students preferred the tailored web site and found it easier to use.
5. CONCLUSIONS
This paper describes an adaptive hypermedia prototype that produces a tailored informational website for applications in a Public Administration domain and its evaluation. It has reviewed the architecture of the application, detailed the forms of the tailoring used and the motivation for using them, and has concluded with a description of two usability studies of the application, the first qualitative and the second quantitative. The qualitative study suggested the desirability of tailored brochures and the acceptability of the use of a query wizard. The quantitative study demonstrated that readers made significantly fewer mistakes with the tailored website and that the use of a detailed query wizard did not slow them down, nor did it affect their confidence in their answers.

The next steps for this work are to pursue more detailed evaluations of the effectiveness of tailored information, including the automatically generated query wizard, and to continue work on the authoring tool to enable a public servant to enter the required information appropriately in the prototype.

6. ACKNOWLEDGEMENTS
This research has been funded by the CSIRO-Centrelink Human Services Delivery Research Alliance (HSDRA) and by Calvin College. The authors would like to thank Carol Taylor and Liz McElachlan from the Centrelink Student Communication Team, Matt Barden from the Centrelink Student Payments Team and Susie Hardy and Rowena Alder who ran the study in the Centrelink Concept Lab. Thanks also go to Adam Strickland, who built significant portions of the prototype, to all the students who participated in the experiment and to the anonymous reviewers.

7. REFERENCES


