Strathprints Institutional Repository

Contextual information and assessor characteristics in complex question answering

Cindy Azzopardi, Leif Azzopardi, Mark Baillie, Ralf Bierig, Emma Nicol, Ian Ruthven
and Simon Sweeney
University of Strathclyde
Glasgow
G1 1XH
i-lab@cis.strath.ac.uk

1 Introduction

The ciqa track investigates the role of interaction in answering complex questions: questions that relate two or more entities by some specified relationship. In our submission to the first ciqa track we were interested in the interplay between groups of variables: variables describing the question creators, the questions asked and the presentation of answers to the questions.

We used two interaction forms – html questionnaires completed before answer assessment – to gain contextual information from the answer assessors to better understand *what* factors influence assessors when judging retrieved answers to complex questions.

Our results indicate the importance of understanding the assessor's personal relationship to the question – their existing topical knowledge for example – and also the presentation of the answers – contextual information about the answer to aid in the assessment of the answer.

2 Variables under study

In our participation in this year's track we studied three groups of variables and selected relationships between these variables. The three groups of variables focus on variables relating to the assessors themselves (section 2.1.1), the questions set (section 2.1.2) and the answers presented (section 2.1.3). In this section we discuss the variables we selected and how we measured them. The majority of variables were investigated through the use of interaction forms – html forms presented to and completed by the answer assessors. Each form was required to be answered within three minutes.

We designed two interaction forms: Interaction Form 1 gathered information on the assessor and on the question, Interaction Form 2 gathered information on sample answers to the questions set by the assessors.

2.1.1 Assessor variables

The first group of variables relate to the assessors themselves. In ciqa the same person who assesses the answers to the question also completes the interaction forms. Knowing more about this person could provide useful information regarding preferences for answer formats or how personal characteristics affect the answer assessment process.

To gain information on the assessor we asked for four responses gathered in Interaction Form 1.

- **topical knowledge**. Topical knowledge has been shown to be one of the major factors in assessing relevance [1-4]. Consequently, we asked the assessor to rate their topical knowledge ("*How much do you think you know about the topics in this question*:") of the major topics in the question on a three-valued scale "*not much/same as most people/know quite a lot*").

- **confidence in assessing answers**. Although the questions are created by the assessors themselves this does not guarantee that they will find the task of assessing answers easy. In a previous study [5] we found that asking assessors to rate their confidence in assessing retrieved material was a useful question

in identifying searcher behaviour regarding the assessment process. Consequently, we asked the assessors to rate their confidence in assessing correct answers to their own questions ("*For this question, how confident are you that you could recognise **correct** answers to this question?*") on a three-valued category ("*very confident/depends on the answers returned/not very confident*").

- **prior expectations**. Next we asked the assessor if they already had an expectation of an answer to the question set, i.e. a page of answers. We asked "*Do you have an answer in mind for this question (could you provide an answer without searching)?*" and solicited responses on a four-category scale ("*yes/no/I could provide a partial answer/no answer but have an idea of what an answer might look like*"). The questions set by the assessors are original questions so we expect they have some exposure to the topics in their questions even if they may know little about the topics. The prior expectation question was designed to elicit some information about their knowledge of possible answers, as opposed to the topical knowledge question which elicited information about their knowledge of the topic of the question. The latter two possible responses ("*I could provide a partial answer/no answer but have an idea of what an answer might look like")* were intended to differentiate between situations where the assessor is confident enough to provide a partial answer (but may require more information to provide a full answer) and situations where the assessor does not know the answer but has an expectation of the likely answers or at least the direction of the answer. For example, for topic 32 "*The analyst is especially interested in opinions of scientists as to whether there is a family link between dinosaurs and birds, and what evidence they cite concerning their opinions*", the assessor may suspect that there is a family link between dinosaurs and birds and is looking for confirmatory evidence.

- **variety of opinions**. The final question we asked the assessor was about what they would prefer in terms of a *set* of answers. We asked the assessor about what type of answer set they required -"*For this question, would good set of answers contain?*" – and asked them to respond using one of two options "*a variety of similar opinions (or evidence)/as many different opinions as possible*". Although few topics gave clues to which of these two options would be applicable to the topic, we felt that assessors may well have reasons for asking for the information and may prefer answers of particular direction.

2.1.2 Question variables

The second group of variables relate to the questions and question descriptions before presentation of answers. In studying the questions we are interested in five variables:

- **time**. The questions used in ciqa often relate to news events and the time of these events can be important in detecting good answers. For each question the assessor was asked about preferences on the date of good answers "*For this question, would good answers come from?*" The response was modelled as a categorical variable ("*recent articles/older articles/any articles*"). This variable was asked in Interaction Form 1.

- **number of entities**. The questions in ciqa relate a number of entities via relationships. The number of entities, as marked by the question creators, is a further variable. Number of entities is a categorical variable with values of 2 (e.g. "*What effect does [aspirin] have on [coronary heart disease]?*") or 3 **(**e.g. **"***What [financial relationships] exist between [the United States] and [supporters of the Irish Republican movement]*").

- **type of relationship**. ciqa questions are modelled on templates as shown in Figure 1 with 6 questions from each template. Each question could therefore be assigned to one of the categories: transport/relationship/influence/positional/evidence.

What evidence is there for transport of [goods] from [entity] to [entity]?
What [relationship][1] exist between [entity] and [entity]?
What influence/effect do(es) [entity] have on/in [entity]?
What is the position of [entity] with respect to [issue]?
Is there evidence to support the involvement of [entity] in [event/entity]?

Figure 1: ciqa templates

Question: What financial relationships exist between drug companies and universities?
Description: The analyst is concerned about universities which do research on medical subjects slanting their findings, especially concerning drugs, towards drug companies which have provided money to the universities.

Figure 2: ciqa topic number 32

- **predicted difficulty**. A final variable was the predicted difficulty of the search. Although the assessors were being asked to judge only retrieved answers to their own questions, rather than perform a search themselves, we felt it would be useful to ask for their opinion on how difficult the information problem (i.e. the description field) would be to answer using a general purpose search engine such as Google or MSN. We ask this in case the assessors have any pre-conceived view on the information problem that might affect their judgement of the quality of the answers presented in the second interaction form. That is, if the assessor believes the question is easy to answer they might rate answers more strictly than if they believe the question to be difficult to answer. The predicted difficulty was measured using a 5 point scale ("*very difficult/fairly difficult/cannot predict how difficult/fairly easy/very easy*") to the question "*If you were searching for answers to this question using a web search engine such as Google, how easy do you think it would be to find good answers?*"). This variable was asked in Interaction Form 1.

- **complexity**– the questions set are designed to be complex in the sense that they relate several entities or concepts (e.g. *drug companies* and *universities* in Figure 2). The description field describes why the information is required and other details of the information need promoting the question (e.g. "*the analyst is concerned about universities which do research on medical subjects slanting their findings, especially concerning drugs, towards drug companies which have provided money to the universities*") from question 32 in Figure 2. The description field, therefore, provides additional information that will be used to assess the answers returned by participating groups. For some questions this additional information simplifies the original question, e.g. "*Specifically, the analyst seeks evidence that smugglers use the island of San Andres for such a purpose*"; for other questions the description field extends the original question to include additional questions. For example for question 32 the question asks "*What familial ties exist between dinosaurs and birds?*" whereas the narrative makes it clear that an answer should contain both opinions on whether dinosaurs and birds are related *and* the evidence for such positions ("*The analyst is especially interested in opinions of scientists as to whether there is a family link between dinosaurs and birds, **and** what evidence they cite concerning their opinions*"). A question with several sub-questions that require answer we deemed as being more complex to answer. Conflating complexity with number of facets we performed an internal classification of the number of facets in each topic description. Three internal assessors were asked, independently, to count the number of facets in each topic description. For all except 6 topics the assessors agreed on the number of facets contained in the description. A short discussion resolved the disagreement on these 6 topics to give a number of facets for each topic[2].

[1] where [relationship] is an element of {"financial relationships", "organizational ties", "familial ties", "common interests"}

[2] Topic 31 was agreed to have 3 facets, topics 26, 27, 29, 30, 33, 34, 35, 38, 41, 42, 44, 45, 46, 48, 49, 50, and 51 had 2 facets and the remaining topics had only 1 facet. We found no correlation between the number of entities in the question and number of facets in the topic description.

2.1.3 Answer variables

Interaction Form 2 presents a series of 8 answers to the assessor. In our study we did not attempt any novel question answering research. Our major interest was in the presentation of answers and the effect of contextual information in the assessment process, in this case date, source and quality information. All presented answers were selected from manual searching of the web. Although the answers to be returned to ciqa for assessment were to be answers from the AQUAINT news collection we felt that we could obtain better answers from the general web. All answers contain a textual answer with ellipses to denote missing text if the answer is a fragment of a sentence e.g. "*..an appearance on Oprah or Today can shoot book sales through the roof…*". We deliberately selected short answers, rather than whole sentences or paragraphs, to simulate the main question answering task in which short answers are preferred. Thus, what we were trying to do was investigate interaction with a good questions answering system.

Each answer had a common layout consisting of three offwhite[3] fields comprising the answer and contextual information and three pale yellow fields containing our questions to the assessor regarding the answer. The first answer line contained the answer to the question presented in red font, the second line contained a source for the answer presented as a URL and the date of the article presented in a dark blue, and the third line contained sources that supported (agreed) with the answer.

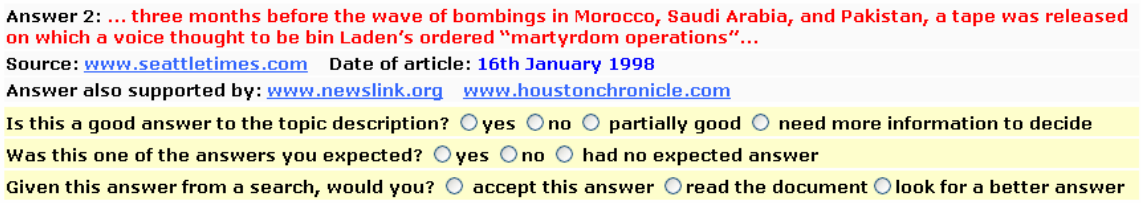

Figure 3: Answer template in Interaction Form 2

When presenting the answers we set out to investigate three variables

- **time of answer**. Answers were either presented with information on the date of the source containing the answer. This was to test whether the date of the information was useful in assessing the answer**.** Sources were randomly assigned dates from one of two lists of dates: recent dates, in this case only from 2006, or older dates, in this case prior to 2004.

- **quality of source.** Each answer was associated with a source which was a website URL. All URLs shown were presented as hypertext links but were not linked to any other page, i.e. clicking on the text did not transfer the assessor to a new page. Although all answers presented were genuine, the sources were manually assigned and did not correspond to the actual sources of the answer. Rather we sought to distinguish between high and low quality sources of information. High quality sources of information were ones that we felt that most assessors would recognise as established sources of reputable information, even if they did not agree with any particular political stance or editorial policy of these sources. We developed a list of these sources which were primarily chosen from a list of top US newspapers and several well known television stations. Low quality sources of information were ones that we felt assessors would be unfamiliar with, primarily sources that had unusual names. The answers provided bear no relation to the actual content of these sources; the sources were only used to test whether the source of the information was important in assessing the quality of an answer presented on the interaction form.

- **supporting evidence.** According to Barry and Schamber [6] one of the important criteria in assessing relevance is the presence of supporting or confirmatory evidence. That is, evidence that information (in our case an answer) is supported by multiple sources can lead to the information

[3] The colour of these fields may not appear very strong in this paper version. After initial pilot testing we reached a balance between contrast of information and visual separation of answers to ciqa questions (offwhite) and questions to the assessor (yellow)

being more likely to be assessed relevant. Accordingly we presented some answers as having multiple sources of information agreeing on the answer. For example in Figure 3 the answer is given by www.seattletimes.com and supported by www.newslink.org and www.houstonchronicle.com. If an answer had supporting sites these correspond to the perceived quality of the original source, i.e. high-quality sources were supported by high-quality sources and weak-quality sources were supported by weak-quality sources. It would have been useful to mix these two conditions (quality of original source vs quality of supporting sources) but the number of combinations required would have been too many to assess within the three minute condition. These supporting sources were also manually assigned and bear no relationship to the actual content of the sources. This allowed us to test whether supported answers were preferred.

The cross combination of three variables (recent vs older information, high-quality source vs low-quality source, supporting vs no supporting sources) gives 8 combinations of answer presentation.

For each answer we asked the assessor to assess:

- **quality of answer**. The first question, asked on all answers, was on the general quality of the answer "*Is this a good answer to the topic description*" and assessors were asked to respond using the categories "*yes/no/partially good/need more information to decide*". "*Partially good*" was intended to reflect answers that supply some useful information but not necessarily all the required information, and "*need more information to decide*" to reflect the situation where the assessor would need more context from the document to decide on the value of the answer.

- **expectation of answer**. Next we asked about the fit of the answer to the assessors prior expectation of the answers -"*was this one of the answers you expected"* which was to be answered using the categories "*yes/no/had no expected answer"*

- **next action**. Finally we asked what the assessor would do given this answer from a search "*Given this answer from a search would you?*". In this case the answers were limited to "*accept this answer/read the document/look for a better answer*".

A final set of questions, displayed after all answers, asked the assessor about the set of answers as a whole, Figure 4. We first asked whether the set of answers provided useful information. The answers themselves might answer the assessors need without reading the full text of the documents ("*yes*"), or might be inappropriate answers ("*no*") or the assessor may require to read the documents to judge how useful the answers were ("*depends on the actual documents*"). Next we asked what, if anything, would have made the answers more useful. Here we had three choices and the assessor could select any combination. Answers could have been more useful if they were longer, more varied (i.e. contained more different types of information or different answers) or more complete (i.e. contained more facets of the initial question and description). Finally we asked what the assessor would do given this form (set of answers) from a search, either browse the documents themselves or start a new search. As we mentioned previously we were interested in the use of answers and surrogates in web search. A poor set of answers might lead to a new query whereas a good set of answers should encourage the searcher to explore the documents retrieved. This final question was intended to reflect an overall assessment of the answers.

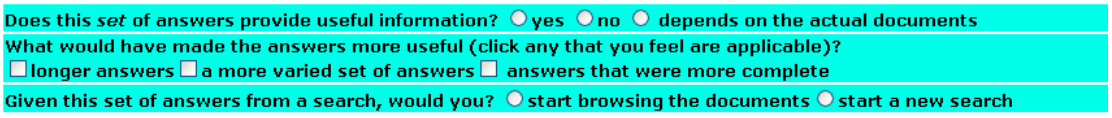

Figure 4: Questions on quality of answer set

3 Results

In this section we present some initial findings starting with general trends. Firstly, examining the quality of the answers provided in the forms as classified by the ciqa assessors in Table 1 we can see that the majority of answers (57%) were deemed to be good answers to the questions and a far smaller percentage

were rated as being poor answers (16%). For a small percentage of answers (slightly more than one answer per form on average) the assessor could not decide on the quality of the answer without reference to the entire document. That is, the answer on its own did not allow the assessor to make a decision with seeing the answer in the context of the entire article. A small number of questions were rated as only partially good.

Response to question "Is this a good answer to the question?"	**percentage**
yes	57.48%
no	16.36%
need more information to decide	15.42%
partially good	10.75%

Table 1: Percentage of answers in each assessment category

Secondly, as can be seen from Table 2 the most common responses from the assessors was to assert an average level of topical knowledge with a high confidence in their ability to assess the accuracy of answers to the questions even though the questions were perceived to be difficult to answer. Even though the assessors felt that they only had average knowledge of the topics in the questions for most questions (22 out of 30) they had sufficient information to guess at least a partial answer to the question suggesting a certain degree of existing topical knowledge. The date of answers in most cases was perceived not to be important for these questions but for most questions the preferred output was a set of different answers rather than very similar answers.

Questions	**Responses**				
knowledge of major topics in question	know a lot n=5	**same as most n=17**	not much n=8		
confidence in assessment of answers	**very confident n=22**	depends on the answers returned n=8			
prior expectation of answers	yes n=4	no answer but have an idea of what an answer might look like n=7	**I could provide a partial answer n=15**	no n=4	
variety of opinions required	**as many different opinions as possible n=20**	a variety of similar opinions (or evidence) n=7			
time of relevant information	recent articles n=3	older articles n=4	**any articles n=23**		
predicted difficulty	very easy n=0	fairly easy n=3	cannot predict how difficult n=9	**fairly difficult n=16**	very difficult n=3

Table 2: Assessors' responses to Interaction Form 2. Most common response shown in bold

3.1 Answer variables

Regarding the presentation of answers we had three main variables: date of answer, quality of answer source and presence/absence of supporting sources.

In Table 3 we look at what percentage of answers in each assessment category were labelled as coming from good or poor sources, e.g. 50.6% of good answers were presented as coming from a source we felt would be recognised as a good source of information whereas 49.4% of good answers came from poor sources of information. Good answers were equally as likely to come from good sources as weak sources. Poor answers, however, were more likely to be rated as coming from weak sources and partially good answers from good sources. Answers therefore may benefit from having good sources but be weakened by coming from weak source.

When examining the next action based on good/weak sources, Table 4, we see that there is a very slight tendency to read documents from good sources whereas the far more common response for answers from weak sources is to look for a better answer.

	good answers	**poor answers**	**cannot decide**	**partially good answers**
good sources	50.6	43.8	49.0	57.8
weak sources	49.4	56.2	51.0	42.2

Table 3: Percentages of answers rated under different categories.

	read	**accept**	**move**
good sources	52.7	50.8	42.6
weak sources	47.3	49.2	57.4

Table 4: Percentages of actions rated under different categories.

In Table 5 we repeat the analysis for the presence/absence of supporting information. Here there is s a slight trend in that good answers more often supported and poor answers are far more likely not to be supported. Unsupported answers also are more likely to be seen as only partially good. The presence of supporting evidence causes more answers to be judged as cannot decide. Perhaps, although we cannot check from this data, the presence of supporting evidence for a poor answer leads to some uncertainty in whether the answer actually is poor. When examining the next action based on good/weak sources, Table 5, we see that there is a slight tendency to read documents from supported sources, a more marked tendency simply to accept the answers without reference to the document for supported answers whereas the more common response for unsupported answers is to look for a better answer.

	good answers	**bad answers**	**cannot decide**	**partially good answers**
presence of supporting information	52.8	39.8	60.2	36.4
absence of supporting information	47.2	60.2	39.8	63.6

Table 5: Percentages of answers rated under different categories.

	read	**accept**	**move**
presence of supporting information	48.4	54.1	44.5
absence of supporting information	51.6	45.9	55.5

Table 6: Percentages of actions rated under different categories.

Finally in Table 7 we repeat the analysis for the recency of information. Here there is again a slight trend in that good answers more often recent, but the stronger pattern is related to the uncertain categories: older answers lead to more cannot decide category and partially good answers. When examining the next action based on good/weak sources, Table 8, we see that there is a tendency to read older answers (perhaps to verify the information which may be out of date), and to accept recent answers without reading.

	good answers	bad answers	cannot decide	partially good answers
recent	52.6	50.2	43.3	45.3
older	47.4	49.8	56.7	54.7

Table 7: Percentages of answers rated under different categories.

	read	accept	move
recent	44.7	54.7	51.2
older	55.3	45.3	48.8

Table 8: Percentages of actions rated under different categories.

3.2 Assessor variables

We established a number of assessor variables. So far, we only have had time to evaluate the effect of a few of these but we present these to give indications of the importance of recording this class of information. First we compare how topical knowledge affects the judgements on the answers given and the predicted next actions based on the answers. In Table 9, we compare the percentage of answers rated as good/poor/etc under the variables for topical knowledge. For the topics the assessor feels they know little (*not much*) the tendency is to be conservative: relatively low use of the definite categories (good/poor answer) and higher use of the partially good and cannot decide categories. Indeed the majority of answers for low topical knowledge reflect some uncertainty regarding the quality of the answer which requires resolution from the whole document. This is indicated in Table 10 where the most common next action for assessors with low topical knowledge is to decide they would read the whole document.

Assessors with higher levels of topical knowledge (*same as most*, *know a lot*) can be more decisive about the quality of answers presented with at least 85% of answers being rated as good or poor and few cases where the assessor cannot decide on the quality of the answer or rates the answer as being partially good. Assessors with the highest level of topical knowledge are far more likely to act on the answer itself without recourse to the full text as, for 95% of answers, the predicted next action is to either accept the answer as presented or move to find a better answer. For the middle range of topical knowledge (*same as most*) the most likely action is one based solely on the answer (*accept* or *move*) but for almost 40% of answers the assessor would seek further information from the document (*read*).

	not much (n=8)	same as most (n=17)	know a lot (n=5)
average good	32.14%	**71.04%**	**54.86%**
average poor	7.14%	14.78%	34.64%
average cannot decide	**37.50%**	7.84%	4.00%
average partially	23.21%	6.34%	6.50%

Table 9: Knowledge and answer quality. Highest value within each knowledge category shown in bold

	not much (n=8)	same as most (n=17)	know a lot (n=5)
read	**69.64%**	39.68%	5.00%
accept	21.43%	**41.10%**	**54.86%**
move	8.93%	19.22%	40.14%

Table 10: Knowledge and predicted next action. Highest value within each knowledge category shown in bold

Examining the relationship between prior expectation and answers, Table 11, we see that if assessors who have no prior expectation of what answers might look like have a very distinct pattern reflecting a conservative approach to assessment: no answers are rated bad, an almost even split between good and partially good and a high rate of cannot decide decisions. This group of assessors also felt they would read the majority of documents, Table 12. For the other groups the percentage of answers rated good fell and percentage of answers rated as bad increased as prior expectation of answer reduced.

Looking at predicted next actions, Table 12, we see that for assessors with high prior expectation the most common next action was predicted to be simply accepting the answer. For assessors with a partial answer in mind the most common predicted next action was to accept the answer (but close second was to look for a better one). For assessors who had a low expectation of answers the more common action was to read the document.

	yes (n=4)	**no but idea (n=7)**	**partially (n=15)**	**no (n=4)**
average good	**68.75%**	**65.31%**	**52.78%**	33.93%
average bad	6.25%	16.90%	32.64%	0.00%
average cannot decide	12.50%	11.33%	9.72%	**34.38%**
average partially	12.50%	6.45%	4.86%	31.70%

Table 11: Prior expectation and answer quality. Highest value within each expectation category shown in bold

	yes (n=4)	**no but idea (n=7)**	**partially (n=15)**	**no (n=4)**
read	36.88%	**39.17%**	24.17%	**76.79%**
accept	**53.75%**	38.01%	**40.83%**	23.21%
move	9.38%	22.83%	35.00%	0.00%

Table 12: Prior expectation and predicted next action. Highest value within each expectation category shown in bold

3.3 Question variables

We have not finally analysed all of the variables involved in our analysis but here present the same analysis as previously for the relationship type in Tables 13 and 14. Ciqa investigated five types of relation in 2006 (transport, relationship, effect, position and evidence). One striking observation is that for some question types there were seen to be more good answers than others. For example, for the effect and evidence question types, the assessors rated the answers as being 80% good whereas for the relationship type only 24% of answers were seen as good and more answers were seen as bad. Whether this is because good answers are easier to find for some questions (i.e. better answers were presented) or whether the answers were easier to evaluate by the assessors is not something we can answer within the current ciqa protocol but are working on.

We can summarise the answers and next actions for each question type as follows:

- **Transport type.** Here there was a fair proportion of good answers, compared to the average of 58% Table 1, but a relatively high proportion of answers where the assessor could not decide on the quality of the answers. Even though the assessor felt that they would accept 24% of the answers, there still seems to be some uncertainty over the quality of answers in this group as, for most answers, the assessor would read the document containing the answer.
- **Relationship type.** Here answers were not seen as being very good: low proportion of good answers, higher proportion of poor answers and most answers being only partially good or needing more information to decide on their quality. Very few answers were good enough to accept without further information and a high number would be rejected (30.83%) in favour of a search for better answers.

- **Effect type.** This class of question contained one of the highest proportions of good answers and one of the lowest proportion of poor and cannot decide decisions. However, for almost half the answers the assessor would read the document containing the answer.
- **Position type.** This type of question contained a fair number of good answers and poor answers. However the good answers appear to be very good as half of the answers presented would be accepted without further recourse to the document.
- **Evidence type.** This class of answers had the highest proportion of good answers and the lowest proportion of cases where the assessor could not decide on the quality of answer. It also had a high proportion of accept decisions: cases where the assessor would accept the answer as presented.

	transport	**relationship**	**effect**	**position**	**evidence**
average good	**51.79%**	23.61%	**80.24%**	**56.55%**	**80.95%**
average bad	10.71%	**28.47%**	6.55%	22.62%	12.50%
average cannot decide	27.50%	**28.47%**	5.42%	10.42%	2.08%
average partially	10.00%	19.44%	7.80%	10.42%	4.46%

Table 13: Question type and answer quality. Highest value within each expectation category shown in bold

	transport	**relationship**	**effect**	**position**	**evidence**
read	**57.50%**	**57.50%**	**44.52%**	25.00%	22.92%
accept	24.17%	11.67%	40.48%	**50.30%**	**64.58%**
move	18.33%	30.83%	15.00%	24.70%	12.50%

Table 14: Question type and predicted next action. Highest value within each expectation category shown in bold

4 Summary

We have many more analyses to run on the ciqa data from this year's track but what we have uncovered already demonstrates that answer assessment is not a neutral process. As well as question type, section 3.3, having an effect on the answer assessment process, so does the assessor's personal context, section 3.2, and how the answers are presented, section 3.1.

5 References

[1] L. Wen, I. Ruthven, and P. Borlund. The effects on topic familiarity on online search behaviour and use of relevance criteria.. Proceedings of the 28th European Conference in Information Retrieval (ECIR 2006). 2006

[2] I. Hsieh-Yee. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. JASIS. 44. 3. 161–174. 1993

[3] D. Michel. What is Used During Cognitive Processing in Information Retrieval and Library Searching? Eleven Sources of Search Information.JASIS. 45. 7. 498–514. 1994.

[4] S. Serola and P. Vakkari. The anticipated and assessed contribution of information types in references retrieved for preparing a research proposal. JASIST. 56. 4. 373-381. 2005.

[5] I. Ruthven, M. Baillie and D. Elsweiler. The relative effects of knowledge, interest and confidence in assessing relevance. Journal of Documentation. in press.

[6] C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: a cross-situational comparison. Information, Processing and Management. 34. 2/3. pp 219-237. 1998.