[AUTHOR'S COPY]

Factor Analysis of a Search Self-Efficacy Scale

Kathy Brennan, Diane Kelly, Yinglong Zhang
School of Information and Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC, 27599-3360
[kbrennan, diane.kelly, yinglongz] @ unc.edu

ABSTRACT
Participants in information search studies are often asked to characterize their search expertise using questionnaires provided by researchers. These questionnaires often contain ad hoc sets of items in part because there are no valid and reliable measures of search experience. In this paper, we present results of an exploratory factor analysis of 327 responses to a 14-item Search Self-Efficacy scale (SSE) that we have been using as a way to measure search experience in our research. The responses come from eight different interactive information retrieval (IIR) studies, which we have conducted over the last six years, with a variety of participant types: university students, participants from the general adult population and crowd-sourced participants. The purpose of this analysis is to understand the variation in search self-efficacy scores across different types of people and to evaluate the potential of the SSE scale as a tool for measuring search experience. Overall, participants from all eight studies reported similar levels of search self-efficacy; the overall average from all eight studies was 7.47 (items scored on a 10-point scale) with little variance (standard deviation=1.36). Both the lowest and highest scores (7.1 and 7.8) were observed in studies involving the general adult population. A factor analysis showed that the questionnaire items load onto six factors, although only four had sufficient numbers of items loading on them. These four factors represent overall task success, effective use of time, query development skills, and advanced search skills.

Keywords
Search behavior, self-efficacy, search experience, measures

1. INTRODUCTION
In many studies of online search behavior, it is common for researchers to ask participants to provide an indication of their search experience. This is typically done with one or more items that ask searchers to indicate the frequency with which they perform searches, the length of time they have used search engines, or their general expertise (from novice to expert). The basic idea is that differences in search experience may cause differences in search behaviors and outcomes [4, 13, 14].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org
CHIIR'16, March 13-17, 2016, Carborro, NC, USA.
Copyright is held by owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3751-9/16/03$15.00.
DOI: http//dx.doi.org/10.1145/2854946.2855002

Measuring search experience was important in early studies of interactive searching as online information systems often required searchers to have extensive knowledge of search syntax and query languages, and were designed primarily for search experts such as librarians. After search went 'public' in the form of Web search engines, more and more people had an opportunity to search, to develop search skills and, more importantly, to develop ideas about the quality of their skills, even to the point of overestimating their skills [6]. Although it is unlikely that today's average Web searcher's skills are equivalent to the skills of a trained searcher, it is reasonable to assume that the public's general search abilities have increased during the past 20 years and that a greater range of expertise exists. This implies that more nuanced measures of search experience are needed if researchers would like to use search experience as a variable in their studies.

Previous measures used to characterize search experience have been coarse and general, and there are no contemporary measures of search experience that are valid and reliable (at least not demonstratively). Typically, researchers create reasonable sounding self-report instruments on an ad-hoc basis [12], or ask questions that greatly simplify search experience, such as how frequently one conducts online searching. Past measures have also not been particularly predictive of behaviors. For example, when distinguishing between search novices and experts, some studies have found that expert searchers submit longer queries than novice searchers [1, 8], while results from other studies have contradicted these findings [2]. Examples of conflicting results in studies of expert versus novice searchers can also be found related to other search behaviors such as number of queries submitted to the system, number of items opened, and speed of querying (see [13] for a review). Smith [13] suggests that these conflicts occur because expert searchers dynamically adapt their search behaviors to their contexts. Thus, behavioral signals alone do not provide an accurate characterization of search ability and also do not allow a person to be characterized before they commence searching.

Self-report instruments are one of the primary methods for gathering information from people (regardless of field of study), so creating measures that are valid, reliable and discriminating are important concerns. One of the only examples of a validated measure of search experience is the work of Debowski, Wood and Bandura [5], who used Bandura's concept of self-efficacy as the basis for search experience [3]. Bandura posited that the main cognitive drivers of behavior are: (1) beliefs about behavioral outcomes leading to favorable consequences and (2) beliefs about one's ability to perform particular behaviors (self-efficacy). Bandura [3] provides a succinct definition of self-efficacy as "People's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances. It is not concerned with the skills one has but with one's judgments of what one can do with whatever skills one possesses" [p. 391]. Notice the focus is on a person's *beliefs* and

[AUTHOR'S COPY]

perceptions not their *actual* skill. The underlying assumption behind using search self-efficacy as a surrogate for search experience is people who have greater confidence in their abilities to accomplish specific search tasks will be more likely to successfully execute these tasks even if through several attempts.

In Debowski, et al.'s [5] 21-item instrument, a list of items was presented and individuals first indicated whether they could perform the task (yes or no) and then indicated the amount of confidence they felt about their abilities to perform the tasks (10-point scale). Debowski, et al. [5] used classic library science research to identify items for their search self-efficacy scale and performed validity and reliability testing, although it is not described in detail in their paper. For the past six years, we have used a modified version of Debowski, et al.'s scale in our research to characterize participants' search experience. Before we started using the scale, we reviewed the items, deleting some that were no longer relevant (e.g., searching CD-ROMs) and updating others; this reduced the number of items on the scale from 21 to 14. We also changed the format of the response, eliminating the binary question asking participants if they felt they could perform the task described. Since people were asked to indicate how confident they were they could execute the task, we assumed if an individual did not feel they could execute a particular task, they could express this through the confidence scale. The modified version of the SSE scale is shown in the Appendix: Table 1.

While we were able to do some reliability testing in the past, none of our previous studies had sufficient sample sizes for conducting a more substantial inquiry such as can be done with factor analysis. In this paper, we describe results of a factor analysis of the pooled search self-efficacy data collected in eight studies with a combined total of 327 participants. The purpose of this analysis is to understand the variation in search self-efficacy scores across different types of people and to evaluate the potential of the SSE scale as a tool for measuring search experience. Table 2 (Appendix) provides a description of the datasets. All studies used general information search tasks, either using online web search engines or TREC collections with custom search applications.

2. RESULTS

Descriptive statistics (means and standards deviations) for each dataset as well as the statistics for the pooled data are shown in Table 3 (Appendix). Ten participants' data were eliminated because they either did not respond to all of the questions or answered all 14 questions with the highest value. There were no significant differences detected among participants' average search self-efficacy scores according to study, $F(7, 319)=.978$, $p=0.447$, which provides support for combining the datasets. It is also the case that this scale was completed at the start of all studies, before any experimental variables were introduced.

Overall, participants reported similar levels of search self-efficacy; the average from all eight studies was 7.47 with little variance (SD=1.36). Both the lowest and highest scores (7.1 and 7.8) were observed in studies involving the general adult population. The pooled averages for the individual questionnaire items range from 6.41 to 8.17. There were three scores greater than 8.0: identify requirements, distinguish relevant, and competent effective. The lowest scores (<7.0) were on four items: special syntax, like a pro, few irrelevant, and focus query.

2.1 Factor Analysis

Before performing Exploratory Factor Analysis (EFA), we conducted Mayer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of Sphericity to test whether the sample size in our study could support a valid EFA. The result of the KMO was 0.92, with Bartlett's Test yielding 357.88 (p<0.001). Both values indicated that our sample satisfies the requirement assumptions for proceeding with EFA [7].

To determine the number of factors to keep in the EFA, we used a procedure called parallel analysis, which is a Monte Carlo simulation technique for determining how many factors to extract during factor analysis [10]. In parallel analysis, one generates a random dataset with the same number of responses and variables as in the sample data. A correlation matrix is then created and eigenvalues computed. Analyses of the eigenvalues allow researchers to determine the appropriate number of factors to extract. The scree plot with an induced parallel analysis using our data indicated that six factors should be kept for the EFA.

We adopted oblique (Promax) rotation because responses to different items were highly correlated. The principal axis factoring (PAF) method was used to extract factors. Readers will likely be more familiar with principal components analysis (PCA) for factor analysis; however, Hatcher [7] argues that PCA should only be used for data reduction as it has limited capability for discovering underlying factor structures. Hatcher [7] argues that PAF provides a better way to identify underlying latent structures, especially when the assumption of multivariate normality in the manifest variables cannot be strictly met.

The item loadings on the six factors are shown in Table 4. All loadings (highlighted in gray shadow and bold typeface) ranged in value from 0.626 to 0.818. The final communality estimate, h^2, of each item is larger than 0.40, indicating that each item is moderately correlated with its corresponding factor. Unlike PCA, the factor loadings for PAF are not always as distinctive, which is why, for example, Item 7, loads on multiple factors. In PAF, it is recommended in these cases to also examine all factor matrices and consider the logical agreement between items [15]**.** For example, we ultimately grouped Item 7 with Items 3 and 4, rather than Items 14 and 9 because the latter items were about time management, while the former were about advanced search skills and Item 7 asked people to indicate the extent to which they could find results similar to a professional.

Table 4: Principal Axis Factoring of Self-Efficacy Data

Item	F_1	F_2	F_3	F_4	F_5	F_6	h^2
6	**0.818**	0.593	0.488	0.553	0.581	0.460	0.68
5	**0.812**	0.572	0.499	0.586	0.588	0.704	0.83
12	**0.812**	0.742	0.589	0.650	0.765	0.400	0.74
13	**0.795**	0.582	0.519	0.613	0.610	0.324	0.65
14	0.622	**0.954**	0.548	0.601	0.621	0.450	0.62
9	0.591	**0.769**	0.526	0.518	0.582	0.438	0.60
8	0.402	0.403	**0.775**	0.365	0.373	0.499	0.63
10	0.523	0.606	**0.767**	0.492	0.499	0.360	0.61
2	0.666	0.637	0.528	**0.982**	0.595	0.531	0.97
1	0.572	0.599	0.421	**0.677**	0.590	0.447	0.97
11	0.626	0.596	0.479	0.522	**0.902**	0.404	0.83
3	0.329	0.407	0.466	0.405	0.341	**0.626**	0.43
4	0.596	0.532	0.408	0.566	0.613	**0.638**	0.57
7	0.603	0.677	0.633	0.491	0.509	**0.630**	0.62

Note: h^2 = communalities of the measured variables

While the PAF produced a solution using six factors, one factor did not have a sufficient number of items loading on it (F_5). One (F_3) had items loading on it that were similar to another factor (F_4), so we decided to eliminate this factor. We believe the four remaining factors represent: overall task success (F1), effective use of time (F2), query development skills (F4) and advanced search skills

(F6). The Pearson correlations observed among the six factors are shown in Table 5, and ranged from 0.45 to 0.74. All correlations were significant (p<0.001).

Table 5. Pearson Correlations of Six Factors

Factor	F_1	F_2	F_3	F_4	F_5	F_6
F_1	1.00					
F_2	0.72	1.00				
F_3	0.59	0.65	1.00			
F_4	0.69	0.67	0.53	1.00		
F_5	0.74	0.71	0.54	0.66	1.00	
F_6	0.46	0.51	0.50	0.50	0.45	1.00

3. DISCUSSION

Overall, levels of search self-efficacy were similar across the eight studies, with little variance. The descriptive statistics show a positive skew of individual and pooled questionnaire data, which is not surprising, given the known tendency for people to overestimate their search abilities [6]. Since self-efficacy is based on people's beliefs about their abilities to accomplish tasks rather than their actual abilities, such a scale might simply reflect this bias. At the same time, however, there is a clear trend in the data showing that when it comes to more sophisticated search skills such as use of special syntax or skills associated with professional searchers, people are able to recognize and acknowledge their more limited abilities. This is particularly the case with studies involving student participants (S4, S5, S8); despite Gross and Latham's finding [6] that students express inflated self-efficacy, the fact that the lowest scores for the item "like a pro" are from the three student studies suggests that people can differentiate between layperson search skills and professional search skills. It may be the case that more directed questions related to search-specific skills are needed to elicit more accurate self-appraisal from study participants. The positive skew might also reflect the supposition that people are getting better at search, or people's interpretation of the number 7 as average. In regard to the lack of variance, the study reporting the largest variance (S7) had the smallest sample size (N=20) and was of the general adult population. The lack of variance in the remaining studies may be the product of a homogeneous sample or an indicator that the samples were of adequate size to reduce sample size-related variance issues. The lack of variance also is a sign that the items may not be particularly discriminating.

We believe our results and analysis extend the original work of Debowski et al. [5] in several ways. First, Debowski et al. investigated search behaviors in a structured, CD-ROM database of bibliographic records, while our studies have all included evaluation of full text documents and Web searching. Our increased scope of information behaviors to include full text evaluation provides a more complete representation of current-day information searching. Second, in all of our studies, participants were provided search tasks and allowed to search either the open web or a document collection based on their own knowledge of searching and using natural language queries. This is different from Debowski et al. [5], where participants were provided search instructions to conduct searches in a structured system that required the use of defined keywords and search syntax. Experimental manipulations in our studies focused on interfaces, whereas the manipulation in [5] involved the instructions given to participants. Finally, in terms of the factor analysis, we can compare our results to Debowski et al. [5]. Debowski et al. identified four subfunctions of electronic search: problem definition, keyword identification, structured search statement construction, and personal beliefs about performance. It is possible to see how three of our four factors map roughly to the four subfunctions: our overall task success to Debowski et al.'s problem definition, query development skills to both keyword identification and structure search statement construction, and advanced search skills to people's beliefs in their abilities to achieve different levels of search performance. In addition, we identified one new factor related to time.

4. CONCLUSION

We presented results of an exploratory factor analysis of data pooled from eight IIR studies using a 14-item SSE scale. The work makes two main contributions: (1) it provides data describing how a variety of different types of people characterize their search self-efficacy; and (2) it suggests four possible factors which might be used in a measure of search expertise. While these factors represent some activities that might distinguish different levels of search expertise they are not exhaustive and are limited to a particular search context; that is, searching for documents to resolve information-gathering tasks. Our results also showed that the scale in its current form is unlikely to be sensitive enough to discriminate amongst people with different levels of search expertise. Thus, additional work needs to be done to enhance the content and discriminant validity of the instrument. This line of research is important as the development of a valid and reliable measure of search expertise would facilitate more in-depth investigations of the relationship between search experience, search interfaces and search behaviors.

5. ACKNOWLEDGEMENT

Thanks to Henry Feild for sharing the data from S1 and everyone who helped gather the data in S2-S8.

6. REFERENCES

[1] Aula, A. & Käki, M., 2003. Understanding expert search strategies for designing user-friendly search interfaces. In *ICWI 2003*, 759-762.

[2] Aula, A. & Nordhausen, K., 2006. Modeling successful performance in Web searching. *JASIS&T*, *57*, 12, 1678-1693.

[3] Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Englewood Cliffs, N.J.

[4] Bates, M. J., 1990. Where should the person stop and the information search interface start? *IP&M*, *26*, 5, 575-591.

[5] Debowski, S., Wood, R., & Bandura, A., 2001. Impact of guided exploration & enactive exploration on self-regulatory mechanisms & information acquisition through electronic search. *J Appl Psychol*, *86*, 6, 1129-1141.

[6] Gross, M. & Latham, D., 2013. Addressing below proficient information literacy skills: Evaluating efficacy of evidence-based educational intervention. *L&ISR, 35*, 3, 181.

[7] Hatcher, L., 2013. *Advanced Statistics in Research: Reading, Understanding and Writing Up Data Analysis Results*. Shadow Finch Media, LLC.

[8] Hölscher, C. & Strube, G., 2000. Web search behavior of Internet experts and newbies. *Compu Netw, 33*, 1-6, 337-346.

[10] Ledesma, R. & Valero-Mora, P., 2007. Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Pract Assess Res Eval*, *12*, 2, 1-11.

[12] Moore, J. L., et al., 2007. The search experience variable in information behavior research. *JASIS&T*, *58*, 10, 1529-1546.

[13] Smith, C. L., 2015. Domain-independent search expertise: A description of procedural knowledge gained during guided instruction. *JASIS&T*, *66*, 7, 1388-1405.

[14] Wildemuth, B. M., 2004. The effects of domain knowledge on search tactic formulation. *JASIS&T*, *55*, 3, 246-258.
[15] Worthington, R. & Whittaker, T., 2006. Scale development research: A content analysis and recommendations for best practices. *Couns Psych*, *34*, 6, 806-838.

APPENDIX

Table 1. Questionnaire items, labels, and questionnaire language (10-point scale, where 1=Totally Unconfident; 5-6=Reasonably Confident; 10=Very Confident)

Item	Label	Questionnaire language: "How confident are you that you can …"
1	Identify requirements	Identify the major requirements of the search from the initial statement of the topic.
2	Develop queries	Correctly develop search queries to reflect my requirements.
3	Special syntax	Use special syntax in advanced searching (e.g., AND, OR, NOT).
4	Evaluate list	Evaluate the resulting list to monitor the success of my approach.
5	Many results	Develop a search query which will retrieve a large number of appropriate articles.
6	Enough results	Find an adequate number of articles.
7	Like a pro	Find articles similar in quality to those obtained by a professional searcher.
8	Few irrelevant	Devise a query which will result in a very small percentage of irrelevant items on my list.
9	Structure time	Efficiently structure my time to complete the task.
10	Focus query	Develop a focused search query that will retrieve a small number of appropriate articles.
11	Distinguish relevant	Distinguish between relevant and irrelevant articles.
12	Competent effective	Complete the search competently and effectively.
13	Little difficulty	Complete the individual steps of the search with little difficulty.
14	Allocated time	Structure my time effectively so that I will finish the search in the allocated time.

Table 2: Description of eight studies used in analysis[1]

Year	Code	Purpose of study	N	Participant type	Corpus
2010	S1	Search behaviors	100	Crowdworkers	Open web
2012	S2	Search interface study	29	General pop. adults	Open web
2012	S3	User evaluations of queries	40	General pop. adults	Open web
2013	S4	Search interface study	36	University students	TREC 2005 robust
2014	S5	Use of query suggestions	29	University students	TREC 2005 robust
2014	S6	Individual differences and search behaviors	47	General pop. adults	Open web
2014	S7	Individual differences and search behaviors	20	General pop. adults	Open web
2015	S8	Search interface study	36	University students	TREC 2005 robust

Table 3: Descriptive statistics from each of eight studies and overall pooled

	S1	S2	S3	S4	S5	S6	S7	S8	Pooled
Sample size (N)	97	29	39	36	29	46	19	32	327
1.Identify requirements	7.8 (2.0)	8.3 (2.1)	8.1 (1.5)	8.3 (1.2)	8.0 (1.6)	8.7 (1.4)	8.5 (2.4)	8.5 (0.9)	8.17 (1.59)
2.Develop queries	8.1 (1.8)	8.3 (1.6)	7.8 (1.7)	7.6 (1.7)	7.8 (1.4)	8.3 (1.4)	8.2 (2.5)	8.0 (1.8)	7.98 (1.58)
3.Special syntax	6.8 (2.2)	6.8 (2.6)	6.8 (2.4)	6.7 (2.1)	6.6 (2.1)	7.5 (2.1)	5.4 (3.1)	6.6 (2.1)	6.82 (2.29)
4.Evaluate list	7.7 (1.9)	7.8 (2.3)	8.1 (1.6)	7.4 (1.9)	7.8 (1.5)	8.5 (1.5)	7.3 (3.1)	7.9 (1.2)	7.73 (1.88)
5.Many results	8.2 (1.8)	8.0 (1.9)	7.8 (1.9)	7.6 (1.7)	7.7 (1.4)	8.1 (1.5)	7.4 (2.9)	7.9 (1.1)	7.68 (1.84)
6.Enough results	7.3 (2.0)	8.1 (1.5)	7.9 (1.5)	7.8 (1.6)	7.8 (1.6)	8.2 (1.6)	8.2 (2.3)	8.2 (1.1)	7.90 (1.68)
7. Like a pro	8.1 (1.7)	6.3 (2.4)	6.7 (2.1)	6.3 (2.1)	6.5 (1.7)	7.3 (1.6)	6.4 (2.8)	6.0 (1.2)	6.69 (1.99)
8.Few irrelevant	7.7 (1.9)	6.2 (2.0)	6.3 (2.1)	6.5 (1.5)	5.8 (1.7)	6.7 (1.7)	6.0 (2.8)	6.8 (1.4)	6.41 (1.95)
9.Structure time	7.6 (2.0)	7.1 (2.1)	6.9 (1.8)	6.9 (1.7)	7.0 (1.9)	7.3 (1.7)	6.6 (2.9)	7.3 (1.7)	7.21 (1.94)
10.Focus query	6.9 (2.1)	6.5 (2.5)	6.2 (2.1)	6.5 (1.8)	6.7 (1.6)	6.9 (1.8)	5.9 (2.5)	6.6 (1.9)	6.54 (2.00)
11.Distinguish relevant	7.6 (1.9)	8.4 (1.9)	7.9 (1.8)	8.1 (1.5)	7.9 (1.7)	8.2 (1.7)	7.5 (2.8)	8.2 (1.1)	8.04 (1.75)
12.Competent effective	7.2 (2.2)	8.1 (1.7)	7.9 (1.6)	7.8 (1.7)	7.8 (1.7)	8.2 (1.6)	8.2 (2.2)	8.3 (1.1)	8.03 (1.64)
13.Little difficulty	8.0 (1.5)	8.0 (2.0)	7.7 (1.8)	7.8 (1.7)	7.8 (1.5)	8.0 (1.8)	7.9 (2.5)	8.0 (1.3)	7.77 (1.81)
14.Allocated time	7.9 (1.7)	7.3 (2.6)	7.3 (1.7)	6.9 (1.8)	7.8 (1.5)	7.6 (1.7)	6.7 (3.0)	7.6 (1.3)	7.49 (1.93)
Total	7.6 (1.4)	7.5 (1.6)	7.4 (1.4)	7.3 (1.2)	7.3 (1.3)	7.8 (1.3)	7.1 (2.1)	7.5 (1.0)	7.47 (1.36)

[1] List of studies is available online at http://ils.unc.edu/ssestudies