8-2010

A Comparative Analysis of Persona Clustering Methods

Jon Brickey
University of Colorado Denver, Jonalan.Brickey@email.ucdenver.edu

Steven Walczak
University of Colorado Denver, Steven.Walczak@ucdenver.edu

Tony Burgess
Center for the Advancement of Leader Development & Organizational Learning (CALDOL), Tony.Burgess@usma.edu

A Comparative Analysis of Persona Clustering Methods

Jon Brickey
University of Colorado Denver
Jonalan.Brickey@email.ucdenver.edu

Steven Walczak
University of Colorado Denver
Steven.Walczak@ucdenver.edu

Tony Burgess
Center for the Advancement of Leader Development & Organizational Learning (CALDOL)
Tony.Burgess@usma.edu

ABSTRACT

Current and future information systems require a better understanding of the interactions between users and systems in order to improve system use, and ultimately, success. The use of personas as design tools is becoming more widespread as academicians and practitioners discover its benefits. This paper presents an empirical study comparing the performance of existing qualitative and quantitative clustering techniques at the task of identifying personas and grouping system users into those personas. A method based on Factor (Principal Component) Analysis outperforms two others using Latent Semantic Analysis and Multivariate Cluster Analysis.

Keywords

Personas, user-centered design, user interaction, interfaces, community of practice.

INTRODUCTION

The failure of information systems (IS) is well-documented in literature spanning business, psychology, computer science and management information systems literature (Cook, 1996, Davis, 1989, Zhang et al., 2005). There are as many dimensions to the problem as there are IS development and design approaches. Since the beginning of modern computing, organizations continue to struggle with challenges encountered at the nexus of humans and computers as they strive to reap the promised rewards of IS productivity (Gerlach and Kuo, 1991).

Given the importance of information systems in our globally-networked society, it is no wonder researchers and practitioners pay so much attention to system development and design, as well as their impact on system acceptance and use. As organizations rely more on digital capital to drive innovation and achieve competitive advantages in worldwide markets (Sambamurthy et al., 2003), it is more important than ever to understand the antecedents of system success and avoid the causes of failure.

Information system design is a significant problem for organizations because it affects system use and success. Increasingly, system designers are turning to the use of personas—a fairly recent design method—as an effective approach to designing better system interfaces in hopes of improving user satisfaction and system success. As organizations become interested in adopting persona development as a design tool, they are faced with different approaches to persona development and left with many questions about the methodologies. Design teams often collect user data from a number of sources using qualitative and quantitative research methodologies; however, depending on the clustering technique used, researchers often ignore relevant data necessary for identifying and grouping users into clusters that ultimately become the personas.

This paper compares the clustering performance of existing qualitative and quantitative persona development methods. This research attempts to answer the following questions:

1. What are the existing qualitative and quantitative methods used for identifying personas?
2. How do the various quantitative and qualitative methods used in identifying personas differ?
3. How well do existing semi-automated qualitative and quantitative techniques perform when compared to manual clustering?

This paper is organized as follows: it begins with a literature review of personas and an overview of the persona development process. The next section compares qualitative and quantitative methods for identifying personas and grouping information system users into similar groups. We then present an empirical study that compares the persona clustering performance semi-automated methods with expert clustering. The paper concludes with a discussion of this work and recommendations for future research.

LITERATURE REVIEW

Personas

Alan Cooper (1999) leveraged the theoretical underpinnings of user-centered design to create a new design methodology known as goal-directed design. As part of his design methodology, he developed a technique to keep product developers focused on a small group of users known as personas. Personas are fictitious characters that represent real system users; they serve as design tools by keeping system interface designers focused on the needs, goals, and frustrations of users (Cooper, 1999).

The use of personas as design tools is becoming more widespread as academicians and practitioners discover its benefits. Since personas are primarily found in literature written by practitioners, there is little academic research on its application and effectiveness (Long, 2009, Miaskiewicz and Kozar, 2006, Pruitt and Adlin, 2006). Several studies cite a focus on the user as the primary benefit of personas, but they also mention other significant advantages of personas such as better communication among developers (Cooper, 1999, Dharwada et al., 2007, Miaskiewicz et al., 2008), and improved user task performance (Dharwada et al., 2007). Task performance has long been a focus of HCI research; however, more recently, there is greater interest in user perceptions and behaviors in the context of IS (Davis, 2006).

Persona development—conducted by persona development teams (PDTs)—may create personas using purely qualitative or quantitative data; however, they often use both types of data at some point in the persona development life cycle (Drego et al., 2008, Javahery et al., 2007, Mulder and Yaar, 2006, Sinha, 2003). For example, researchers in most of the literature use qualitative data to identify user groups, but they also use quantitative data—such as demographics or level of computing skills—to describe typical persona dimensions. Cooper (1999) assumes designers will use qualitative data to create personas and Goodwin (2002) strongly recommends it. Pruitt and Grudin (2003) support iterative data collection and persona development using qualitative and quantitative methods to improve the "selection, enrichment, and evolution of Personas" (p. 8).

Comparison of Qualitative and Quantitative Persona Clustering Methods

Qualitative and quantitative clustering techniques include manual and semi-automated methods. Manual methods require human judgment to identify users with similar characteristics; semi-automated methods rely on the use of statistical software for analysis. PDTs combine participants (system users) into personas using qualitative or quantitative clustering techniques. The persona grouping technique employed depends on the type of data—qualitative or quantitative—PDTs gather and which technique they prefer. Qualitative research usually provides text and quantitative research usually results in numbers, as depicted in Figure 1. The manual qualitative method of clustering users into personas is prevalent in the persona literature (as shown in Figure 2); however, recent studies include various attempts to automate the process using qualitative and quantitative methods.

Qualitative Persona Clustering Methods

Using qualitative data, researchers usually cluster participants using manual techniques such as affinity diagrams, card sorting exercises, and expert panels (Broschinsky and Baker, 2008, Lindgren et al., 2007). Researchers advocating the manual approach to clustering prefer the use of rich qualitative data from interviews or observations (Cooper et al., 2003, Goodwin, 2002). Pruitt & Adlin (2006) recommend grouping participants by themes and relationships, then determining which ones are most important to the project objectives. Mulder and Yaar (2006) suggests a collaborative, qualitative process using various stakeholders gathered "in a room with plenty of whiteboards" to develop clusters based on goals, usage lifecycle, or a combination of behaviors and attitudes (p. 123).

Although the manual qualitative method of clustering users into personas is popular among PDTs; it has at least two main drawbacks. First, as the number of participants and textual data grows, it becomes difficult for human experts to make objective judgments and trace their findings back to user data (Pruitt and Adlin, 2006). Whereas humans may have qualitative and quantitative data available for making clustering decisions, they tend to focus on qualitative data for insight on a few users at the expense of a broader understanding of users through the analysis of quantitative data (Sinha, 2003). Second, manual clustering methods may require extensive resource commitments from organizations. Development teams require specialized skills in qualitative research methods and it may take considerable time and money to obtain results (Javahery et al., 2007, Miaskiewicz et al., 2008, Sinha, 2003).

Some critics of the manual qualitative persona clustering techniques advocate more automated methods that they believe address the rigor and resource concerns of the manual qualitative method. Although the manual qualitative method of clustering dominates the persona literature, recent efforts to create semi-automated clustering techniques using qualitative or quantitative data have captured the attention of researchers and practitioners (McGinn and Kotamraju, 2008, Miaskiewicz et al., 2008).

Miaskiewicz et al. (2008) recently applied Latent semantic analysis (LSA) to demonstrate the effectiveness of a semi-automated qualitative persona clustering technique. LSA is considered a qualitative method because it requires the collection of textual data, such as interviews, even though the analysis is performed by software and the results are displayed as quantitative data. Derived from theory similar to factor analysis, LSA is defined by Landauer, Foltz and Laham (1998) as "a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text" (p. 2). LSA determines similarity of word passages by only using the contexts in which words appear and do not appear; it has been proven to make judgments similar to humans in standardized tests and essay grading (Landauer et al., 1998). The limitations of the LSA persona development method are that it requires knowledge in specialized research tools and processes, and its ability to compare participants depends on the corpus of text employed in the analysis.

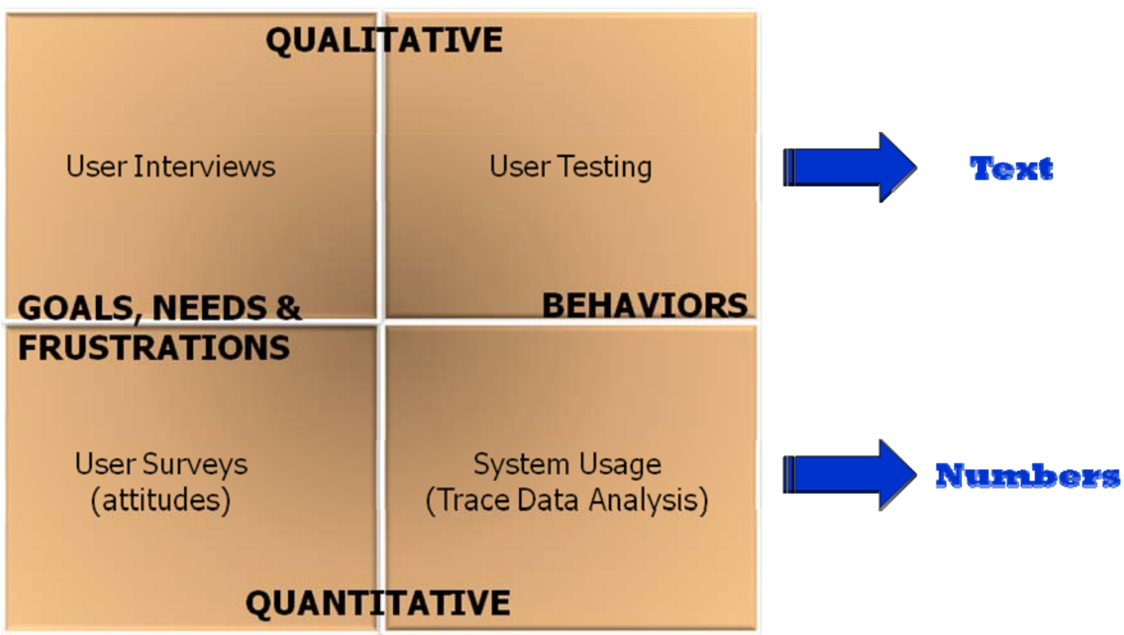

Figure 1. Qualitative Versus Quantitative Data for Persona Development. Adapted from Mulder and Yaar (2006)

Quantitative Persona Clustering Methods

Quantitative clustering techniques are capable of determining the relationships between multiple variables to find patterns in user data that may be latent, or unobservable to the human eye. The quantitative clustering techniques found in the persona literature are all semi-automated methods because they rely on statistical software to identify clusters of users; the techniques include exploratory factor analysis n (McGinn and Kotamraju, 2008), principal component analysis (Sinha, 2003) and multivariate cluster analysis (Javahery et al., 2007). Proponents of quantitative techniques assert they overcome some of the drawbacks of qualitative clustering: subjective assignment decisions and a lack of rigor, the need for experience in qualitative research training, cognitive limitations of humans and considerable resource commitments (Javahery et al., 2007, McGinn and Kotamraju, 2008, Miaskiewicz et al., 2008, Mulder and Yaar, 2006, Sinha, 2003).

Although quantitative clustering techniques rely on statistical software for analysis, PDTs must still collect data, prepare it for analysis, and interpret the results. The quantitative data required for these techniques must be collected as numbers or converted to numbers in order to use statistical software to produce clusters. PDTs gather quantitative data through surveys (measures of goals, needs and frustrations), system transaction logs (usage), or organizational records (demographics, education levels, skills, and other descriptive data). As shown in Figure 2, each of the quantitative clustering techniques can only be found once in the literature; therefore, each technique is explained below in general mathematical terms and as implemented by previous authors. Table 1 summarizes the main characteristics of persona clustering techniques.

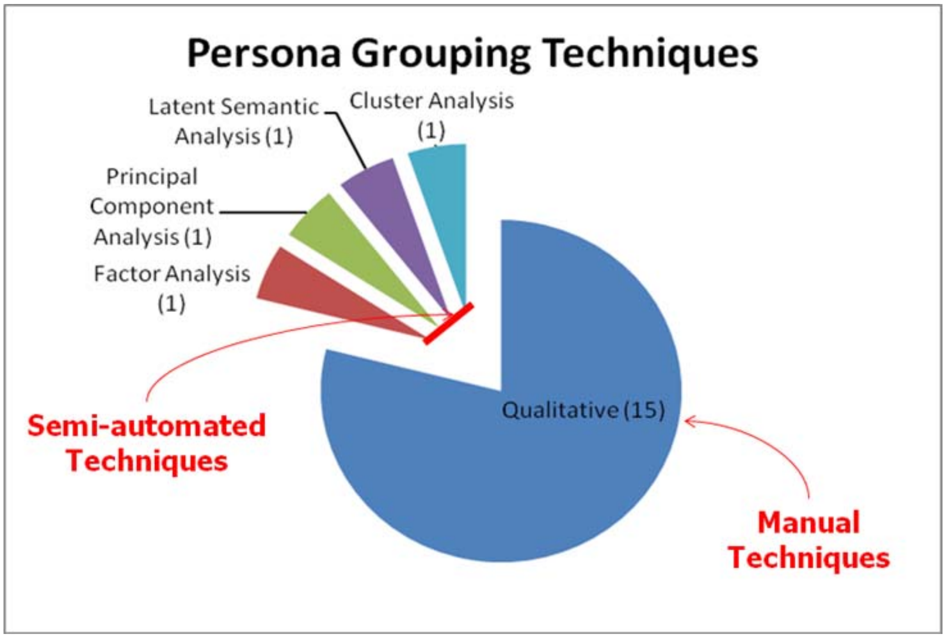

Figure 2. Persona Grouping Techniques Found in 19 Studies (Numbers Represent the Number of Times the Technique is Found in the Literature)

Sinha (2003) was the first to propose a semi-automated quantitative technique for persona clustering. His approach implements Principal Component Analysis (PCA)—a data reduction technique—to identify 3 components (groups of users) for an online restaurant finder based on 32 dimensions of the restaurant experience (Sinha, 2003). PCA reduces the original variables into new components that convey as much of the original data as possible (Morgan et al., 2007). The author claims his PCA approach provides a direct link from user surveys to the identification of persona clusters while nearly automating the task (Sinha, 2003).

Factor Analysis (FA) is a data reduction technique concerned with identifying the latent structure among multivariate data (Hair et al., 2006). McGinn and Kotamraju (2008) apply FA to identify latent groupings of system users based on work tasks and demographics. They claim their FA clustering method is "fast and cheap, easy to scope," and results in "data-driven" clusters derived from statistically significant sample sizes (McGinn and Kotamraju, 2008, p. 1521). The main limitation of FA as applied to persona clustering is that after identifying clusters, PDTs still needed to conduct interviews to validate the clusters and collect qualitative data to explain *why* users behave in ways identified by FA. Depending on the number of variables in the analysis and the communalities among them, FA and PCA may arrive at similar results (Tabachnick and Fidell, 2007). Therefore, in the rest of this paper we treat them as a single method for persona clustering.

Multivariate Cluster Analysis (MCA) has been used in "every research setting imaginable" to group objects (Hair et al., 2006, p. 561). Like FA and PCA, MCA is capable of analyzing large amounts of data and determining relationships between participants that would be difficult, if not impossible, for humans to identify through visual inspection. MCA aims to measure similarity between participants using a variety of algorithms found in statistical software. The researcher determines the number of desired clusters and the software groups participants based on the similarity of their attributes. MCA is often criticized because it will identify clusters based on any input data, even if there is no underlying structure in the data (Hair et al., 2006).

Technique	Approach	Data Type	Benefit	Limitation
Manual	Qualitative	Text	Rich, qualitative data; human judgment	Humans can be overloaded by quantity of data; considered subjective, time consuming, requires expertise
LSA	Qualitative	Text	Human-like judgment; semi-automated; objective; data-driven	Analysis depends of corpus of text used to compare interviews; requires knowledge of LSA tools
FA/PCA	Quantitative	Numeric, interval	Semi-automated; objective; data-driven; identifies latent factors; reduces data complexity	Quality of the analysis depends on the ability to capture desired goals, needs, frustrations and behavior
MCA	Quantitative	Numeric, ratio	Semi-automated; objective; data-driven; multiple algorithm options	Clusters any input data, even if there is no real underlying structure

Table 1. Comparison of Persona Clustering Techniques

Mulder and Yaar (2006) describe how to use pivot tables in Microsoft Excel to sort users based on selected characteristics and perform an MCA-like clustering effect, but his technique does not group users using multivariate data—it merely provides a different way to visualize the data. Javahery et al. (2007) implement a more sophisticated MCA clustering method using statistical software to analyze multivariate data. The authors sample 22 users of a biomedical system and identify 3 clusters based on iterative cluster analysis: the first iteration of MCA clusters users based on domain experience and system use; the second iteration clusters using age as a factor (Javahery et al., 2007). Their technique demonstrates the use of readily available user data—such as skills, experience, system usage and demographics—to quickly collect data and identify clusters. However, their approach does not capture user goals, needs and frustrations.

RESEARCH METHODS

The use of personas is on the rise for the design of information systems, but although Cooper introduced the concept a decade ago, there is little research into the specific development techniques and procedures (Long, 2009). The clustering of IS users into personas is critical to the success of resulting personas because it identifies groups of users sharing similar goals, needs, frustrations and behavior—the characteristics upon which personas are based. To date, there is no attempt in the persona literature to compare the performance of existing clustering techniques within the domain of persona creation. The primary goal of the semi-automated clustering techniques is to reduce the resources required to identify personas: it takes time, money, and training to manually identify clusters (McGinn and Kotamraju, 2008, Miaskiewicz et al., 2008, Sinha, 2003). This research compares the performance of all the existing semi-automated qualitative and quantitative persona clustering techniques found in the literature. Figure 3 depicts the conceptual model for comparing these techniques.

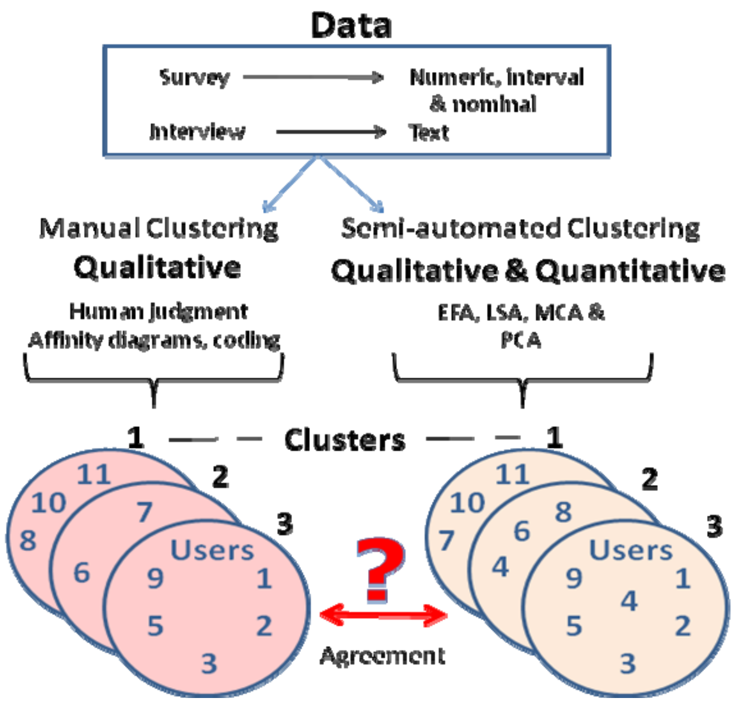

Figure 3. Comparing the Performance of Persona Clustering Methods

Experiment Design

Our empirical study used data gathered through an online survey and server queries of system usage for an online knowledge management system (KMS). The target population was members of Army professional forums, which are a type of KMS similar to distributed communities of practice found in many civilian organizations (Brickey and Walczak, 2010, Dixon et al., 2005). The sampling frame consisted of current, active members in the Army's Company Command forum, which is developed and facilitated by the Army's Center for the Advancement of Leader Development and Organizational Learning (CALDOL). The forum is not only one of the Army's most active and successful, but it is also the test forum that directly influences nearly 40 other forums.

The forum facilitators emailed survey solicitations to 116 captains assigned to the same Army installation; 26 members of the forum initially agreed to participate in the survey; therefore, the response rate was 22%. In the end, however, 6 participants did not complete all sections of the survey, leaving 18 valid responses. The survey consisted of three sections: 10 semi-structured interview questions focused on user goals, needs and frustrations for an online community of practice; 16 rating scale questions focused on the same dimensions; and several questions varying in structure to capture user demographics, experiences and interests (to provide detailed persona characteristics). The semi-structured interview questions (to be used in the LSA method) were developed by adapting questions found in Goodwin (2002), Miaskiewicz et al. (2008), Mulder and Yaar (2006), and Pruitt and Adlin (2006) to meet the needs of the forum. The rating scale questions (to be used in FA/PCA) were adapted from Sinha (2003) and Mulder and Yaar (2006).

The forum administrator also queried the server logs to record each participant's system usage behavior in the form of counts—the number of times members performed certain online activities in the forum, such as page views, document downloads, contributions to discussions, and other activities. The server logs recorded 12 variables for user activity. The validity of using such online transaction log data has been questioned, but when done with care, has been found to be an effective, unobtrusive technique for understanding user behavior (Jansen et al., 2008).

Whereas there is no agreed upon "gold standard" for persona clustering, the manual technique serves as the de-facto baseline for comparing semi-automated techniques (Miaskiewicz et al., 2008). A panel of two experts reviewed the survey responses and the system usage data to create three persona clusters using existing manual qualitative methods. Experts were selected based on significant experience in the fields of user interface design and facilitation of communities of practice. In cases where the two experts did not agree on the clustering of a participant, they resolved the clustering to achieve a consensus assignment.

Performance of the existing semi-automated qualitative and quantitative clustering techniques was measured by comparing the assignment agreement for each method with the manual expert clustering. A simple way to express agreement between the semi-automated and manual clustering techniques is to calculate percentage agreement by counting the number of matching user assignments into clusters and dividing by the total number of users (participants in the study). Whereas percentage agreement is widely used, it is misleading because it does not account for the fact that some percentage of the agreement is expected by chance (Barrett et al., 1990). Cohen (1960) introduced a measure of agreement that adjusts the percentage agreement due to chance. Cohen's kappa, a statistical measure of inter-rater agreement for categorical items, is used in persona literature to compare LSA and expert clusters (Miaskiewicz et al., 2008). Although interpretation of Cohen's kappa is a matter of debate, Landis and Koch (1977) provide a guideline for interpreting degree of agreement ranging from poor (< 0.21) to almost perfect (> 0.80) .

RESULTS

Latent Semantic Analysis was conducted to assess the similarity among survey participants according to their text responses to semi-structured questions. Participant responses were compared using the LSA web tool found at http://lsa.colorado.edu. The resulting cosine matrix—representing the similarity among the interview participants—was analyzed in Statistical Package for Social Sciences (SPSS) version 16 using hierarchical agglomerative clustering. The LSA method resulted in persona clusters of 12, 5, and 1 user. Table 2 displays the cluster assignments for the 18 final participants.

Factor (Principal Component) Analysis with varimax rotation was conducted to reduce 16 variables to 3 components that would serve as personas. Three components were requested in order to make comparisons to the expert clusters. After rotation, the first component accounted for 26.6% of the variance, the second factor accounted for 22.9%, and the third component accounted for 19.7%. Each component consisted of at least three items with a minimum loading of 0.7. Using SPSS, component scores were computed for each of the participants and used to conduct hierarchical agglomerative clustering. The FA/PCA method resulted in clusters of 4, 9, and 5 users as shown in Table 2.

Multivariate Cluster Analysis using the Chi-squared measure for counts was conducted in SPSS to assign participants to 3 clusters according to online behavior as recorded in the 12 variables stored in the system's user logs. Table 2 shows the cluster assignment of 6, 10, and 2 users using the MCA method.

	Persona Clusters		
	1	**2**	**3**
Method	Users Assigned to Clusters		
Expert	1,8,9,13, 16,18	3,5,7,14, 15,20,22, 24	10,11,17, 25
LSA	1,8,9,10, 11,13,15, 17,18,22, 24,25	3,5,7,14, 20	16
FA/PCA	1,9,16,22	5,7,8,13, 14,15,20, 24,25	3,10,11, 17,18
MCA	1,5,11,16, 24,25	3,7,9,13, 14,15,17, 18,20,22	8,10

Table 2. Clustering Assignments for Each Method

Cohen's kappa was computed to determine the degree of agreement between the semi-automated persona clustering methods and expert clustering. The performance of each method, in terms of the raw number of matching cluster assignments and the adjusted kappa measure, is listed in Table 3. The LSA kappa of 0.31 indicates a fair amount of agreement with experts; the FA/PCA kappa of 0.48 indicates moderate agreement; and the MCA kappa of 0.19 indicates poor agreement.

Persona Method	Matching Assignments	Kappa	Degree of Agreement
LSA	10/18	0.31	Fair
FA/PCA	12/18	0.48	Moderate
MCA	9/18	0.19	Poor

Table 3. Grouping Agreement of Semi-automated Methods and Experts

DISCUSSION AND CONCLUSIONS

Although semi-automated clustering methods overcome some of the drawbacks of manual methods, they have limitations as well. All three semi-automated methods of persona clustering presented here used hierarchical agglomerative cluster analysis in the final step. Clustering results are completely dependent on the researchers' selection of survey questions and transaction logs that ultimately quantify user goals, needs, frustrations and behaviors. Statistical software gives you results regardless of the quality of the input data, but the results may not adequately explain similarities or differences between user groups.

Miaskiewicz et al. (2008) implemented an LSA methodology for identifying personas and reported substantial agreement (a kappa value > 0.60) between their LSA-derived clusters and those created by experts. They studied information needs for users of a reference library system; our domain consisted of military officers using a KMS. The results of LSA are dependent on the corpus of documents used to compare text. Therefore, it is possible that if we used a more specific, tailored corpus containing military documents we could have achieved a higher kappa.

The usage data collected only represented the number of times a user performed a certain activity, like viewing a page or downloading files. Whereas downloading a document may be a reasonable indicator of a user's interest or behavior, viewing a page may not be an appropriate indicator of user behavior. Fuller and de Graaff (1996) propose a count of page views is a simple measure of user behavior, but recommend capturing additional user information such as the amount of time spent on pages. Alag (2009) also recommends more explicit forms of user interaction, such as voting or rating of online content.

None of the three methods discussed in this paper achieved substantial agreement with manual qualitative clustering performed by experts. However, with modification and improvements, the semi-automated methods show promise as potentially faster, cheaper, and sufficient methods for identifying groups of information system users and developing personas to improve user interface design.

IMPLICATIONS AND FUTURE RESEARCH

Our goal in this study was to empirically compare existing qualitative and quantitative methods of persona clustering. Based on the results of this study, FA/PCA appears to be a viable semi-automated clustering method that moderately agrees with experts using manual qualitative methods; however, more work is necessary. Future research should focus on additional semi-automated persona clustering methods, especially ones that combine different approaches to take advantage of the benefits of each without inheriting their weaknesses. Additional research may also answer other research questions such as at what level of performance does it make sense to trade clustering performance for development and design resources? In other words, what levels of agreement and cost savings justify semi-automated clustering methods?

ACKNOWLEDGMENTS

We would like to thank the Army's Center for the Advancement of Leader Development and Organizational Learning (CALDOL) at the United States Military Academy for partial support for this research and feedback along the way. Additionally, we appreciate the time and patience of the expert panel members.

REFERENCES

1. Alag, S. (2009) Collective intelligence in action, Manning, Greenwich, CT.
2. Barrett, J. F., Jarvis, G. J., Macdonald, H. N., Buchan, P. C., Tyrrell, S. N. and Lilford, R. J. (1990) Inconsistencies in clinical decision in obstetrics, *Lancet*, 336, 8714, 549-551.
3. Brickey, J. and Walczak, S. (2010) A Comparative Analysis of Professional Forums in the United States Army and Hybrid Communities of Practice in the Civilian Sector, in Ralph H. Sprague, Jr., (Ed.) *Proceedings of the* 43rd *Hawaii International Conference on Systems Sciences*, January 5-8, Koloa, HI, USA.
4. Broschinsky, D. and Baker, L. (2008) Using persona with XP at LANDesk software, in *Agile 2008 Conference Proceedings*, Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ, 543-548.
5. Cohen, J. (1960) A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20, 37-46.
6. Cook, M. A. (1996) Building enterprise information architectures: reengineering information systems, Prentice Hall, Upper Saddle River, NJ.
7. Cooper, A. (1999) The inmates are running the asylum, SAMS, Indianapolis, IN.
8. Cooper, A., Reimann, R. and (2003) About Face 2.0: The Essentials of Interaction Design, Wiley Publishing, Indianapolis, IN.
9. Davis, F. D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 13, 3, 319-340.
10. Davis, F. D. (2006) On the Relationship Between HCI and Technology Acceptance Research, in Dennis Galletta (Ed.) Human-Computer Interaction and Management Information Systems: Applications, M.E. Sharpe, New York, 395-401.
11. Dharwada, P., Greenstein, J., Gramopadhye, A. and Davis, S. (2007) A Case Study on Use of Personas in Design and Development of an Audit Management System, *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 469-473.
12. Dixon, N. M., Allen, N., Burgess, T., Kilner, P. and Schweitzer, S. (2005) Company Command: Unleashing the Power of the Army Profession, The Center for the Advancement of Leader Development and Organizational Learning, West Point, NY.
13. Drego, V. L., Temkin, B. D. and McInnes, A. (2008) Use Personas To Design For Engagement, Forrester Research, http://frstweb001.forrester.com/ Research/Document/Excerpt.html.
14. Fuller, R. and de Graaff, J. J. (1996) Measuring user motivation from server log files, in *Proceedings of the Human Factors and the Web 2 Conference*.
15. Gerlach, J. H. and Kuo, F. Y. (1991) Understanding human-computer interaction for information systems design, *MIS Quarterly*, 15, 4, 527-549.
16. Goodwin, K. (2002) Getting from Research to Personas: Harnessing the Power of Data, in *Cooper Newsletter* http://www.cooper.com/content/insights/newsletters/2002_11/ getting_from_research_to_personas.asp.
17. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and Tatham, R. L. (2006) Multivariate data analysis, Pearson Education, Upper Saddle River, NJ.
18. Jansen, B. J., Taksa, I. and Spink, A. (2008) Research and Methodological Foundations of Transaction Log Analysis, in B. Jansen, B. J. Taksa, and A. Spink (Eds.) *Handbook of Research on Web Log Analysis*, Information Science Reference, Hershey, PA.
19. Javahery, H., Deichman, A., Seffah, A. and Radhakrishnan, T. (2007) Incorporating human experiences into the design process of a visualization tool: A case study from bioinformatics, in *IEEE International Conference on Systems, Man and Cybernetics*, Institute of Electrical and Electronics Engineers Inc., Montreal, 1517-1523.

20. Landauer, T. K., Foltz, P. W. and Laham, D. (1998) An introduction to latent semantic analysis, *Discourse processes,* 25, 259-284.
21. Landis, J. R. and Koch, G. G. (1977) The measurement of observer agreement for categorical data, *Biometrics,* 33, 1, 159-174.
22. Lindgren, A., Fang, C., Amdahl, P. and Chaikiat, P. (2007) Using personas and scenarios as an interface design tool for advanced driver assistance systems, in *Universal Access in Human-Computer Interaction, Ambient Interaction,* 4555 Springer Verlag, Heidelberg, Germany, 460-469.
23. Long, F. (2009) Real or Imaginary: The effectiveness of using personas in product design, in *Irish Ergonomics Society Annual Conference,* Dublin.
24. McGinn, J. and Kotamraju, N. (2008) Data-driven persona development, in *The Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems* ACM, Florence, Italy.
25. Miaskiewicz, T. and Kozar, K. (2006) The Use of the Delphi Method to Determine the Benefits of the Personas Method – An Approach to Systems Design, in *Proceedings of the Fifth Annual Workshop on HCI Research in MIS,* December 9, Milwaukee, WI, Association for Information Systems, 50-54.
26. Miaskiewicz, T., Sumner, T. and Kozar, K. A. (2008) A latent semantic analysis methodology for the identification and creation of personas, in *Proceedings of the Twenty-sixth annual SIGCHI conference on Human factors in computing systems,* April 5-10, Florence, Italy, ACM Press, 1501-1510.
27. Morgan, G. A., Leech, N. L., Gloeckner, G. W. and Barrett, K. C. (2007) SPSS for introductory statistics: Use and interpretation, Lawrence Erlbaum Associates, Inc., Mahwah, NJ.
28. Mulder, S. and Yaar, Z. (2006) The user is always right: a practical guide to creating and using personas for the web, New Riders Publishing, Thousand Oaks, CA.
29. Pruitt, J. and Adlin, T. (2006) The persona lifecycle: keeping people in mind throughout product design, Morgan Kaufmann, San Francisco.
30. Pruitt, J. and Grudin, J. (2003) Personas: practice and theory, in *Proceedings of the 2003 conference on Designing for user experiences,* ACM Press, San Francisco.
31. Sambamurthy, V., Bharadwaj, A. and Grover, V. (2003) Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firm, *MIS Quarterly,* 27, 2, 237-263.
32. Sinha, R. (2003) Persona development for information-rich domains, in *Proceedings of Computer Human Interaction '03 Extended Abstracts on Human Factors in Computing Systems,* April 5-10, Ft. Lauderdale, FL, ACM Press, 830-831.
33. Tabachnick, B. G. and Fidell, L. S. (2007) Using multivariate statistics, Pearson Education, Inc., Boston, MA.
34. Zhang, P., Carrey, J., Te'eni, D. and Tremaine, M. (2005) Incorporating HCI Development into SDLC: A Methodology, *Communications of the AIS (CAIS),* 15, 512-543.