



Participatory Educational Research (PER)
Vol. 8(4), pp. 426-438, December 2021
Available online at <http://www.perjournal.com>
ISSN: 2148-6123
<http://dx.doi.org/10.17275/per.21.98.8.4>

Id: 862471

Bibliometric Analysis of Articles on Computerized Adaptive Testing

Meltem Yurtcu*

Department of Educational Sciences, Inonu University, Malatya, Turkey
ORCID: 0000-0003-3303-5093

Cem Oktay Güzeller

Faculty of Tourism, Akdeniz University, Antalya, Turkey
ORCID: 0000-0002-2700-3565

Article history

Received:
16.01.2021

Received in revised form:
05.05.2021

Accepted:
17.05.2021

Key words:

Bibliometric analysis,
Computerized adaptive testing,
Co-citation,
Detecting and Visualizing

The items that are suitable for everyone's own ability level with the support of computer programs instead of paper and pencil tests may help students to reach more accurate results. Computer adaptive tests (CAT), which are developed based on certain assumptions in this direction, are to create an optimum test for every person taking the exam. It then becomes essential to examine the development process of such important exams and to monitor what studies have contributed to this development in what year. Citespace is a program developed to map information fields, explain the relationship between different disciplines, examine and estimate the studies in a certain period of time, uncover the latest studies and predict the trend issues that occur according to the analysis of bibliographic records of related publications. In this study, it is aimed to find out what articles about CAT are produced in which areas, at what time periods and which articles have a significant effect in these periods. CiteSpace program was used to make a document/article co-citation analysis. Articles on CAT between 1946-2016 were scanned by "or" connector. A total of 637 articles were used, the analyses were finalized according to the networks. As a result of the research, clusters were determined based on the relationship in the citations, articles that were the most cited and important among studies on CAT were presented.

Introduction

The measurement and evaluation process can be handled in three classes as diagnostic, formative and summative according to their purposes (Crisp, 2007). Different measurement and evaluation tools are used in all of these processes. On the other hand, it can be said that tests are tools that are widely used to achieve the goals of measurement and evaluation processes. Generally, tests comprise of paper-pencil tests during the in class measurement and evaluation process. However, paper and pencil tests may not always provide the desired level of quality. For example, determining a student's ability in a difficult test or making inferences about the psychometric properties of a test administered to a successful group may not always be obtained correctly. In addition, it may not be economical to apply the paper-pencil test for everyone at the same time and with the same number of questions. Such negative situations

* Correspondency: meltem.yurtcu@gmail.com

may cause misinterpretations and false inferences about individuals or the test. In this respect, directing the items that are suitable for everyone's own ability level with the support of computer programs instead of paper and pencil tests may help students to reach more accurate results.

There are basically two approaches to the development of tests. These approaches are Classical Test Theory (CTT) and Item Response Theory (IRT) (Hambleton, & Jones, 1993; Thompson & Weiss, 2011). The two approaches have advantages over the purposes they are used for. In the classical test theory, tests are developed depending on the group and applied to each student at the same difficulty and discrimination level. However, the difference in the levels of students, which is the most important element of education, should be taken into consideration. In tests developed according to the IRT approach, the abilities of the individuals can be determined independently from the items asked, and information about the items of the test can be obtained regardless of the success of the group (Hambleton, Swaminathan, & Rogers, 1991).

With the development of computer technology, calculations of IRT approaches have been included in computer systems and computerized adaptive testing (CAT) have begun to be used (Meijer & Nering, 1999; Thompson & Weiss, 2011; Weiss & Kingsbury, 1984). With CAT applications, questions with appropriate psychometric characteristics are directed to determine the ability level of each individual (Wise, et al., 1992). If the individual's ability is determined, the exam is completed. In this case, the exam will be of different length for each individual (Meijer & Nering, 1999). Thus, it will be possible to make a more qualified measurement with fewer items (Meijer & Nering, 1999; Wainer, 1993). However, the development or evaluation of CAT requires large sample and cost. CAT applications have more advantages than disadvantages (Meijer & Nering, 1999). Taking advantage of such benefits of CAT applications, CAT studies are applied in many areas. Studies on CAT in measurement and evaluation in education can be gathered under the following areas: in the comparison of item selection methods (Barrada, et al., 2010; Deng, Ansley & Chang, 2010; Han, 2010, 2012; Lee & Dodd, 2012; Sulak, 2013; Veldkamp, 2010); paper-pencil test results in comparison (Kalender, 2011; Kezer, 2013; Smits, Cuijpers & van Straten, 2011); in comparison of the termination rules (Babcock & Weiss, 2012; Choi, Grady & Dodd, 2011; Eroğlu, 2013; Yao, 2013); determining the item pool characteristics (He, Diao & Hauser, 2014; Lee & Dodd, 2012; van der Linden & Xiong, 2013); in the study of differential item functioning (Gierl, Lai & Li; 2013; González-Betanzos, Abad, & Barrada, 2014; Piromsombat, 2014) and to investigate the relationship between cognitive diagnosis models (Cheng, 2009; Huebner, 2010; Hsu & Wang, 2015). In addition to such studies, the inclusion of a study showing for what disciplines and time periods the studies about CAT in the literature have been designed and conducted will give an idea to many researchers and help them to direct their fields of study. It indeed plays a very important role to examine the development process of such important exams and to monitor which of the studies in the relevant literature of the field have contributed to this development in which years.

Therefore, in this study, it is aimed to reveal the most common study areas of CAT, the most common time periods and the most important articles of the specified period of time. This study will shed light on the studies related to interdisciplinary CAT applications and hence is believed to contribute to the literature.

Method

The aim of the research is to classify the articles on CAT and its applications and to reveal the network structure between these articles. In addition, it is aimed to reveal which article is most popular in the specific time line. For this purpose, it has been designed as a Bibliometric research to detect quantitative measurements and indicators. Bibliometric studies are used to compare research on numerous areas (Besimoğlu, 2015), to evaluate and follow scientific processes (Gmür, 2003; Mongeon & Paul-Hus, 2016; Santos, 2015; Van Raan, 2005). These intend to unearth the relationships between documents and examine the development of a research topic with co-citation methods (Tsay, Xu & Wu, 2003; Yu, Chang & Yu, 2016).

The Data of The Research

There are three databases representing different approaches which are Web of Science (WoS), Scopus and Google Scholar. WoS and Scopus are commercial databases and are used as a database to provide current data by evaluating citations and articles (Feng, Zhang, Du & Wang, 2015; Jasco, 2005; Seyedghorban, Jekanyika-Matanda & LaPlaca, 2015). Google Scholar has been an open source since 2004 (Jacso, 2005). Scopus is built from records extracted from Elsevier such as Geobase, Biobase, Embase, and enriched with citation information (Agapiou & Lysandrou, 2015; Archambault, et al., 2009; Fingerman, 2006). WoS is interpreted as a much more scientific and comprehensive multidisciplinary content research platform than Scopus (Fingerman, 2006; <http://thomsonreuters.com/thomson-reuters-web-of-science/>).

The data required for the research were obtained from the Web of Science™ core collection database and from the articles covered by the SCI-EXPANDED, SSCI, A & HCI, CPCI-S, CPCI-SSH, ESCI indices. In the study, the terms “computerized adaptive testing”, “computerized adaptive exams”, “computer adaptive testing”, “computer adaptive test” and “computer adaptive exams” were scanned by “or” connector for between the years 1946-2016. Articles solely on the subject under consideration were discussed in this study. A total of 800 articles were obtained from the database. Repeated articles are excluded from analysis, analyses were continued with 637 articles.

Analysis of Data

In Bibliometric research, there are usually three types of co-citation analysis work which are analysis of journal co-citation, document co-citation and author co-citation. The basic assumption behind co-citation analysis is to see the relevant document is cited from the successful work done in the subject area (Tsay, et al., 2003). If two documents or authors appear in the same bibliography (used source), there is co-citation. The more the two publications are cited together, based on the similarity of the content of these two authors or documents / articles, the stronger their links (Feng, et al., 2015; Gmür, 2003; Tsay, et al., 2003). In this study, document co-citation is used.

CiteSpace is a java application that analyses and visualizes the large network structure obtained for bibliometric research (Chen, 2006; Feng, et. al. 2015; Zhao & Wang, 2011). The program, developed by Chaomei Chen, produces co-citations or C networks of nodes and links. It is an effective program for measuring relationships and links between sources such as authors, articles, institutes, terms and keywords (Tsay, et al., 2003; Seyedghorban, et al., 2015; Zhao & Wang, 2011). In fact, it constitutes a program developed to map information fields, explain the relationship between different disciplines, examine and estimate the studies



in a certain time period, uncover the most recent studies and use these to predict the trend issues that arise according to the analysis of the bibliographic records of related publications (Chen, 2014; Feng, et al., 2015; Khan & Niazi, 2017; Liu, Yin, Liu & Dunford, 2015; Zhao & Wang, 2011). The present study was carried out with the analysis of 637 articles between 1984 and 2016 with CiteSpace program.

Clusters are formed according to the similarities of the references cited by published articles in the feature of interest in CiteSpace program. There are three different algorithms to name cluster. These algorithms are TF*IDF, LLR and MI algorithms. Algorithms serve to characterize the nature of the cluster to be identified (Chen, 2014). The program uses TF*IDF as default. LLR is based on the log-likelihood ratio, while the MI algorithm uses common knowledge (Chen, 2014). In this study, the naming of the clusters was formed based on the words in the abstracts of the articles according to TF*IDF (term frequency by inverted document frequency).

With CiteSpace program, the structural development of important studies in time periods can be observed. Timeline visualization can be used to view new trends and developmental schema (Kim & Chen, 2015; Santos, 2015). In this respect, the time tunnel of the studies about CAT was included in the research.

Findings

The articles obtained from WoS in the research vary between 1984 and 2016. The 30 disciplines where the highest number of articles on CAT were carried out and the information showing how many articles were published in this field are given Table 1.

Table 1. The disciplines where the highest number of articles on CAT are grouped

Discipline	f	Discipline	f	Discipline	f
Psychology	306	Neurosciences	26	Geriatrics Gerontology	5
Mathematical Methods in Social Sciences	138	Neurology		Pharmacology Pharmacy	5
Health Care Sciences Services	100	Psychiatry	25	Oncology	4
Educational Research	71	Business Economics	13	Operations Research Management Science	4
Public Environmental Occupational Health	66	Rheumatology	12	Research Experimental Medicine	4
Rehabilitation	66	Surgery	11	Mathematical Computational Biology	3
Mathematics	64	Engineering	10	Substance Abuse	3
Orthopaedics	42	General Internal Medicine	10	Urology Nephrology	3
Sports Sciences	42	Medical Informatics	9	Anaesthesiology	2
Computer Science	27	Paediatrics	8	Environmental Sciences Ecology	2
		Science Technology Other Topics	6		

When Table 1 is examined, it is observed that the disciplines for which the highest number of articles were produced about CAT are psychology, mathematical methods and health. However, it is seen that studies are carried out in a wide variety of disciplines. It is observed

that CAT is used more in disciplines where it is important to evaluate the individual's level independently from the group.

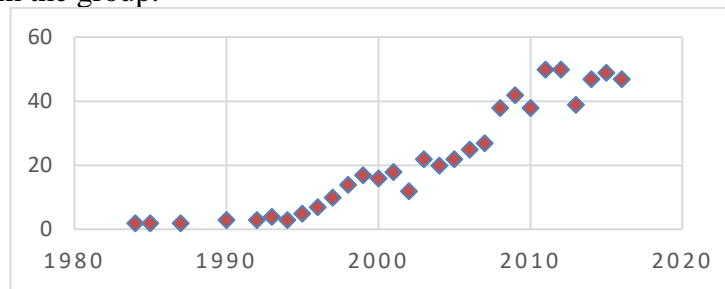


Figure 1. The studies on CAT according to years

Figure 1 shows that the studies on CAT have increased over time. The main reason for this situation can be expressed as the development and spread of technology.

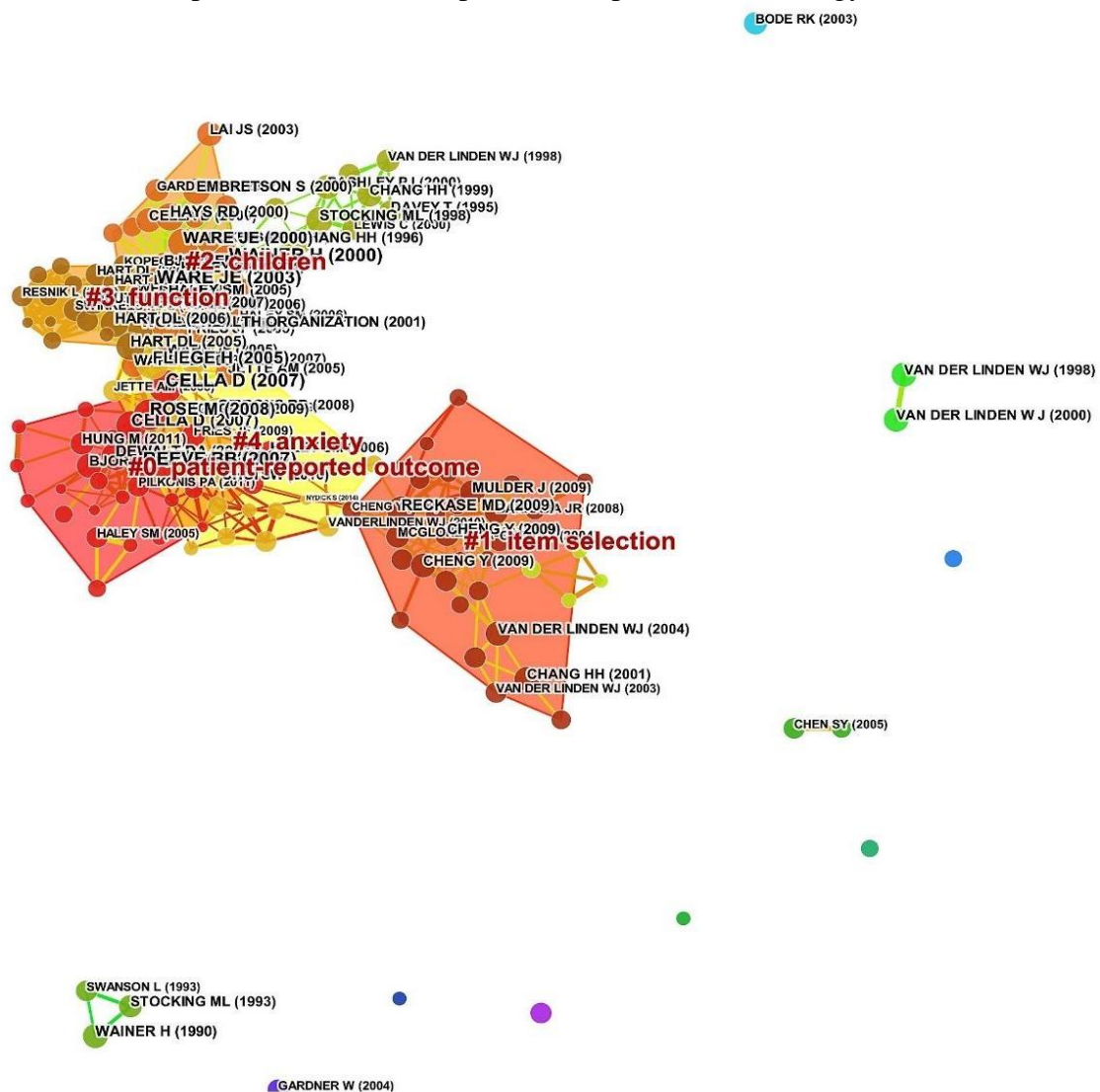


Figure 2. Clusters based on article co-citation network structure

Figure 2 is obtained to visually observe the relationship between the articles to citations. A cluster view of a network of article co-citation is presented in Figure 2. Each node represents an article about CAT and is labelled with the author's name and publication year. Each link

between nodes shows the co-citation relationship between the two articles. There is a total of 166 nodes in Figure 2. Nodes and citation networks vary according to their colour and size. The size of the nodes is proportional to the number of citations. It shows a communication link between the two peaks in the networks. The thickness / thinness of the lines indicates the strength of co-authoring. CiteSpace program provides information with the colours of time periods. The blue colour shows the first years, the green colour shows the middle years, and the orange and red colour show the current years. Darker shadows of the same colours represent earlier time periods, and lighter colours show later times (Khan & e Niazi, 2017). As shown in Figure 2, studies representing large nodes such as Ware (2000), Hart DL. (2005), Cella D. et al (2007) show that they have more citations than other articles. The most cited articles and the information related to these articles are given in Table 2.

Table 2. The most cited articles and information on these articles

Citation Counts	Articles	Cluster #
44	Cella D, Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al., 2007, Medical Care, V45,	0
43	Reeve BB et al, 2007, Medical Care, V45,	0
41	Ware, J. E., Jr., Kosinski, M., Bjorner, J. B., Bayliss, M. S., Batenhorst, A., Dahlof, C. G., et al., 2003. <i>Quality of Life Research</i> , 12(8), P935	2
36	Rose M, Bjorner, J.B., Becker, J., Fries, J.F. & Ware, J. E. , 2008, Journal of Clinical Epidemiology, V61, P17	0
36	Fliege H, Becker, J., Walter O. B., Bjirner J. B., Klapp B. F. & Rose M., 2005, Quality of Life Research, V14, P2277	4
30	Cella D, Riley W., Stone A., Rothrock N., Reeve B., et al. 2010, Journal of Clinical Epidemiology, V63, P1179	0
30	Cella D, Gershon R., Lai J-S. & Choi S., 2007, Quality of Life Research, V16, P133	0
30	Ware, J. E., Bjorner, J. B., Jr., & Kosinski, M., 2000, Medical Care, V38, P73	2
26	Hart DL, Mioduski J. E. & Stratford P. W., 2005, Journal of Clinical Epidemiology, 58(6), 629–638	3

As seen in Table 2, the most cited articles are mainly in the first cluster. The most cited article is Cella D. et al.'s (2007), entitled "The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group during its first two years". In this article, the researchers summarized the organization and scientific activity of the PROMIS network during its first two years.

In the study, six clusters were obtained and the naming of the clusters was formed based on the words in the abstracts of the articles according to TF*IDF (term frequency by inverted document frequency). The size and colours of each cluster differ from the other clusters. The cluster # 0 is the largest cluster. In addition, articles at cluster # 0 with articles at cluster # 4 are highly interrelated.

As a result of the clustering process, there are two coefficients showing the importance of the network obtained from the analysis of 637 articles that include the concept of CAT. These coefficients are "Silhouette" and "Modularity Q". Silhouette values for six clusters ranged from 0.697 to 0.969. A value of 0.3719 was obtained for the mean Silhouette value, and 0.6585 for Modularity Q. These two values are expected to be higher than 0.5 as a good network structure indicator. The modularity Q value is high and this value gives information about whether the articles in the network are logically divided into clusters. The mean Silhouette value shows the homogeneity of the clusters. A high Silhouette value indicates that the cluster members are more stable. However, if the size of the cluster is small, it does not mean that the cluster is homogeneous. For example, when there are only 7 elements in the

cluster # 9, and the Silhouette value is 1, this can mean that the same author can also refer to 7 articles (Chen, 2014).

Table 3. Cluster names and properties determined by article co citation analysis

Cluster ID	Size	Silhouette	Mean (Year)	Label (TFI*DF)	Most cited article to the cluster
0	31	0.841	2009	(13.51) patient-reported outcome; (11.98) pain; (11.71) promis.	Amtmann,, D , Cook KF, Jensen MP, et al (2010) Development of a promis item bank to measure pain interference
1	30	0.919	2007	(10.84) item selection; (10.73) method; (10.2) computer.	Barrada,, JR , Olea J., Ponsoda V. & Abad F. A. (2010).A metod for the comparison of item selection rules in computerized adaptive testing
2	27	0.697	2003	(10.86) children; (10.33) computer; (10.06) adaptive test.	Wang,, Y. C., Hart D. L., Cook K. F. & Mioduski J. E. (2010). Translating shoulder computerized adaptive testing generated outcome measures into clinical practice
3	20	0.881	2006	(11.87) function; (11.42) responsive measure; (10.73) functional status outcome.	Hart,, DL , Werneke M.W, Wang Y-C, Stratford P.W. & Mioduski J.E. (2010) Computerized adaptive test for patients with lumbar spine impairments produced valid and responsive measures of function
4	20	0.746	2009	(11.64) anxiety; (11.27) depression; (11.11) mood.	Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms
5	9	0.969	1998	(9) adaptive test; (8.64) computer; (8.64) computerized adaptive testing.	Hau,, K.T. & Chang, H. H. (2001) item selection in computerized adaptive testing: should more discriminating items be used first?

As shown in Table 3, the largest cluster, whose Silhouette value is 0.841, is cluster # 0 with 31 articles. The average of the publication dates of the works in this cluster is 2009. This cluster is referred to as the “patient-reported outcome” according to TF*IDF. The most cited study is “Development of a PROMIS item bank to measure pain interference” by Amtmann D, Cook KF, Jensen MP, et al (2010).

The second cluster is #1, which is called “item selection” according to TF*IDF. This cluster contains 30 articles. The most cited study in this cluster is “A method for the comparison of item selection rules in computerized adaptive testing” by Barrada, J.R., Olea, J., Ponsoda, V. & Abad, F.J. on 2010. In Table 3, it is possible to see the information about the other clusters. The methodological development of CAT applications is mostly observed in clusters # 1 and # 5, while in other clusters CAT applications in the field of health come to the fore.

Timeline visualization is examined for 6 clusters. Each node in the timelines represent an important article. Rings and colours of articles give information about “betweenness” centrality, citation frequency or citation “burstiness” (Khan & Niazi, 2017). The articles that are important in this directory stand out with their rings. The size of the node is proportional

to the number of citations. The purple nodes are an indication of the centrality betweenness which indicates that they are an important turning point. The citation burstiness is shown in red (Khan & Niazi, 2017).

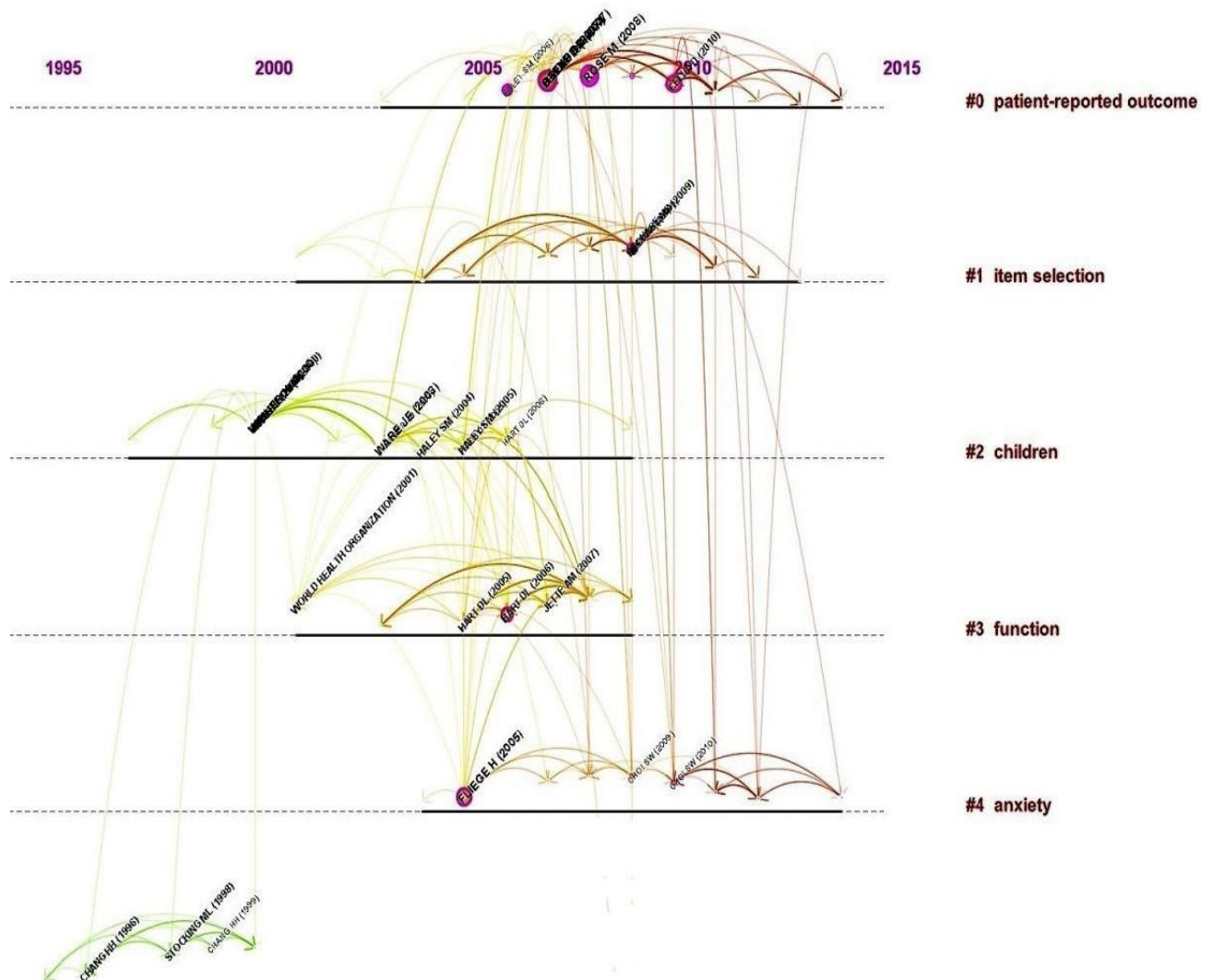


Figure 3. Co-citation timeline for the 6 clusters

The developmental scheme of each cluster was given separately according to the timelines. As some of the important articles in the clusters stand out, the relationship of these studies with each other is shown by networks. The articles that have high betweenness centralities are given below:

Table 4. The articles that have high betweenness centralities

Centrality	Articles	cluster #
0.40	Rose M, Bjorner, J.B., Becker, J., Fries, J.F. & Ware, J. E. (2008). <i>Journal of Clinical Epidemiology</i> , 61,17	0
0.33	Cheng Y, Chang H.-H & Guo, F.,(2009). <i>Educational and Psychological Measurement</i> , 69, 35	1
0.27	Haley S.M., Ni P., Ludlow L.H. & Fragala-Pinkham M. A., (2006), <i>Archives of Physical Medicine And Rehabilitation</i> , 87, 1223	0
0.24	Cheng Y & Chang H.-H, (2009), <i>British Journal of Mathematical and Statistical Psychology</i> , 62, 369	1

0.19	Fliege H, Becker, J., Walter O. B., Bjirner J. B., Klapp B. F. & Rose M., (2005). <i>Quality of Life Research</i> , 14, 2277	4
0.19	Choi SW, Reise S. P., Pilkonis P. A., Hays R. D. & Cella D.,(2010). <i>Quality of Life Research</i> , 19, 125	4
0.19	Reckase MD, (2009), <i>Statistics for Social and Behavioral Sciences</i> , SC, 1	1
0.18	Cella D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al., (2007), <i>Medical Care</i> , 45,	0
0.17	Hart DL, Mioduski J. E., Werneke M. W. & Stratford P. W., (2006). <i>Journal of Clinical Epidemiology</i> , 59(9), 947–956	3
0.16	Reeve BB et al, (2007), <i>Medical Care</i> , 45,	0

These articles in Table 4 represent important turning points in the clusters to which they belong. It is seen that the most important turning points were between 2005 and 2010. In some years, more than one study constitutes a turning point.

As shown in Figure 3, the absence of red rings around the nodes indicates that there is no citation burstiness. Therefore, the citation numbers of the articles did not show a sudden increase in a short time.

As seen from the timeline, articles conducted after 2010 are mostly about reporting on the health of the patient and anxiety and are related to each other. The articles related to item selection have improved after 2010 and it is possible to say that further studies on these issues will be made.

Discussion and Conclusion

Article co-citation analysis is a statistical method used to analyse the structure underneath a prominent topic and to reveal the citations and attributes of the articles (Tsay, et al., 2003; Yu, et al., 2016). It is also used to visualize scientific research, identify emerging content, and predict future research (Song, Zhang & Dong, 2016). The current study is planned to see the structures in articles with CAT applications and to learn what areas include more articles of this sort and how these articles are related to each other. Using 637 articles related to CAT applications between 1984 and 2016, this study draws attention to some important points visually. Article co-citation analysis was performed by using CiteSpace program. Subsequently, the results and some suggestions can be listed as follows:

It is seen that the articles were studied after 1995. and that these articles increased more intensively after 2010. Technological developments and advances in the uses of technology in this sense may be effective in this direction.

CATs have been used in health, education and psychology. However, it is possible to mention that CAT applications are mostly used in health areas. As a matter of fact, it has been observed that the most cited articles in the field of health belong to articles such as Cella et al.'s (2007) and Reeve et al.'s (2007). When the content of these articles is examined, it is figured out that CAT applications in the field of health were used to define the psychometric properties of the individual in determining the pain threshold and to examine the patients' report outputs.

In addition, articles have been carried out to determine the best method selection and to compare these methods. As it can be seen from the results, the use of CAT applications in more than one field will help to enrich interdisciplinary studies and help researchers to see the different perspectives relating to their research.



As a result of the citation analysis, six significant clusters were reached. These clusters contained articles from different fields and in this case, it appeared that it is encouraging to work among disciplines. Therefore, if the words to be used in the analysis are chosen to address more than one area, researchers can conduct a more comprehensive research. This means, researchers will be more active in the fields or journals on the subject of their study and have the opportunity to apply the methods in different fields.

Timelines are provided to acquire an idea of future research and to see the relationship between the obtained 6 clusters. In this timeline, it can be said that the studies on CAT are extremely related to each other. In particular, it is possible to see that the studies in cluster # 0, which includes intensive studies in the field of health, are intensively related to the studies in the four clusters following it and that, the articles in clusters # 2 and # 3 are intensively referring to each other. The articles, which have been displayed with purple colour, indicate that there are important turning points of those years. Therefore, it shows when the important articles have appeared. The most important turning point spotted in the present study is about the research that evaluated the property of item bank with CAT for PROMIS. This study is deemed important for many other related studies.

It is possible to estimate new studies in the field of CATs with the timeline and to see the past studies that help in the development of these studies. CAT applications were first studied with the articles in cluster # 5, referred to as “Adaptive Test”, and then used in clinical studies in the 2000s.

With the development of technology, computer applications have been included in many fields. CAT applications developed in accordance with the Item Response Theory that can be used to determine each individual's level independently from the group. This is very important for education and health areas since it is vital to evaluate each individual differently for these fields. Therefore, the use of CAT applications especially in these areas will provide more qualified results. This research casts light on the work of scholars doing research in the field of CATs, and will be able to reach the information about what studies were completed for what time period in the field of study. For these reasons, bibliometric studies will remain a crucial instrument in order to see what the deficiencies in any topic / field and thusly carry out studies in order to eliminate these detected deficiencies. Researchers can conduct their studies using this analysis in particular while carrying out literature reviews on the field of interest.

References

- Agapiou, A. & Lysandrou, V. (2015). Remote sensing archaeology: tracking and mapping evolution in European scientific literature from 1999 to 2015. *Journal of Archaeological Science*, 4, 192–200
- Archambault, E., Campbell, D., Gingras, Y. & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of The American Society For Information Science And Technology*, 60(7):1320–1326.
- Besimoğlu, C. (2015). *Agricultural Research Trends of Agriculture Faculties in Turkey: Bibliometric Analysis of 1996-2011*. (Unpublished doctoral thesis), Hacettepe University, Ankara.
- Babcock, B. & Weiss, D.J., (2012). Termination criteria in computerized adaptive tests: do variable-length CAT's provide efficient and effective measurement? *International Association for Computerized Adaptive Testing*, 1, 1-18.
- Barrada, J. R., Olea, J., Ponsoda, V. & Abad, F., J. (2010), A method for the comparison of item selection rules in computerized adaptive testing, *Applied Psychological Measurement* 34(6) 438–452.

- Cella, D., Yount, S. Rothrock, N., Gerson, R., Cook, K., Reeve, B., Ader, D., Fries, J. F., Bruce, B., Rose, M. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH roadmap cooperative group during its first 2 years. *Medical Care*, 45, 3-11
- Chen C,(2006). CiteSpaceII. Detecting and visualizing emerging trend sandtransient patterns inscientific literature. *Journal of The American Society For Information Science And Technology*, 57(3):359–377
- Chen, C. (2014). *The CiteSpace manual*. Retrieved from <http://cluster.ischool.drexel.edu/~cchen/citespace>.
- Cheng, Y. (2009). Computerized adaptive testing for cognitive diagnosis. In D. J. Weiss (Ed.), Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing. Retrieved from www.psych.umn.edu/psylabs/CATCentral/
- Choi, S. W., Grady, M.W. & Dodd, B.G.(2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement*, 71, 37-53.
- Crisp, G. (2007). *The e-Assessment Handbook*. Continuum International Publishing Group, London.
- Deng, H., Ansley, T. & Chang, H. (2010). Stratified and maximum information item selection procedures in computer adaptive testing. *Journal of Educational Measurement*, 47(2), 202-226.
- Eroğlu, M. G. (2013). *Comparison of different test termination rules in terms of measurement precision and test length in computerized adaptive testing* (Doctoral dissertation), Hacettepe University, Ankara,
- Feng, F., Zhang, L., Du,Y. & Wang, W. (2015). Visualization and quantitative study in bibliographicdatabases: A case in the field of university–industry cooperation. *Journal of Informetrics* 9, 118–134
- Fingerman, S. (2006). Web of Science and Scopus: Current features and capabilities. *Issues in Science and Technology Librarianship*, 48(Fall). Retrieved from <http://www.istl.org/06-fall/electronic2.html>
- Gierl, M. J., Lai, H. & Li, J. (2013): Identifying differential item functioning in multi-stage computer adaptive testing, *Educational Research and Evaluation: An International Journal on Theory and Practice*, 19(2-3), 188-203.
- Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, 57(1), 27-57.
- González-Betanzos, F., Abad, F. J. & Barrada, J. R. (2014). Fixed item parameter calibration for assessing differential item functioning in computerized adaptive tests. *Psicológica*, 35, 331-359.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47. <https://doi.org/10.1111/j.1745-3992.1993.tb00543.x>
- Han, K. (2010). *Comparison of Non-Fisher Information Item Selection Criteria in Fixed Length Computerized Adaptive Testing*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Denver.
- Han, K. (2012). SimulCAT: Windows Application That Simulates Computerized Adaptive Test Administration. *Applied Psychological Measurement*, 36(1).64-66
- He, W., Diao, Q. & Hauser, C. (2014). A Comparison of Four Item-Selection Methods for Severely Constrained CATs. *Educational and Psychological Measurement*, 74(4), 677-696. doi: doi 10.1177/0013164413517503



- Huebner, A. (2010). An Overview of Recent Developments in Cognitive Diagnostic Computer Adaptive Assessments. *Practical Assessment, Research & Evaluation*, 15(3).1-7.
- Hsu, C.L. & Wang, W.C. (2015). Variabl-length computerized adaptive testing using the higher order DINA model. *Journal of Educational Measurment*, 52(2), 125-143. <http://thomsonreuters.com/thomson-reuters-web-of-science/>.
- Jacso, P. (2005). As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537-1547.
- Kalender, İ. (2011). *Effects of different computerized adaptive testing strategies on recovery of ability*. (Unpublished Doctoral Dissertation). Middle East Technical University, Ankara.
- Khan, B. S. & Niazi, M. A. (2017). Network community detection: A review and visual survey. CoRR, abs/1708.00977.
- Kezer, F. (2013). *Comparison of the computerized adaptive testing strategies*. (Doctoral Dissertation). Ankara University, Ankara
- Kim, M. C & Chen, C. (2015). A scientometric review of emerging trends and newdevelopments in recommendation systems. *Scientometrics*, 104, 239–263
- Lee, H. Y., & Dodd, B. G. (2012). Comparison of exposure controls, item pool characteristic, and population distributions for cat using the partial credit model. *Educational and Psychological Measurement*, 72(1), 159- 175. doi: 10.1177/0013164411411296
- Liu, Z., Yin, Y., Liu, W. & Dunford, M. (2015). Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis. *Scientometrics*, 103,135–158, doi:[10.1007/s11192-014-1517-y](https://doi.org/10.1007/s11192-014-1517-y)
- Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: overview and introduction. *Applied Psychological Measurement*, 23(3), 187–194. doi:10.1177/01466219922031310
- Mongeon, P. & Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228. doi:10.1007/s11192-015-1765-5.
- Piromsombat, C. (2014). *Differential item functioning in computerized adaptive testing: can CAT self-adjust enough?* the University of Minnesota Digital Conservancy, Retrieved from: <http://hdl.handle.net/11299/163281>.
- Thompson, N. A. & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research & Evaluation*. 16(1), 1-9
- Tsay, M.Y., Xu, H. & Wu, C.W. (2003). Author co-citation analysis of semiconductor literatüre. *Scientometrics*, 58(3), 529-545.
- Seyedghorban, Z., Jekanyika-matanda, M. & LaPlaca, P. (2015). Advancing theory and knowledge in the business-to-business branding literature, *Journal of Business Research*. 69(8), 2664-2677, doi:[10.1016/j.jbusres.2015.11.002](https://doi.org/10.1016/j.jbusres.2015.11.002)
- Santos, A. B. (2015). Open Innovation research: trends and influences – a bibliometric analysis. *Journal of Innovation Management*, 3(2), 131-165
- Sulak, S. (2013). *Comparision of item selection methods in computerized adaptive testing* (Doctoral dissertation), Hacettepe University, Ankara.
- Smits, N., Cuijpers, P. & van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research* 188, 147–155.
- Van Raan, A.F.J.,(2005). For your citations only? Hot topics in bibliometric analysis. *Measurement: Interdisciplinary Research and Perspectives* 3 (1), 50–62.
- van der Linden, W. J. & Xiong, X. (2013). Speededness and adaptive testing. *Journal of Educational and Behavioral Statistics*, 38, 418-438.

- Veldkamp, B. P. (2010). Bayesian item selection in constrained adaptive testing using shadow tests. *Psicologica*, 31(1), 149-169
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Applied Psychological Measurement*, 37, 3-23.
- Yu, Y.C., Chang, S. H. & Yu, L.C. (2016). An academic trend in STEM education from bibliometric and co-citation method. *International Journal of Information and Education Technology*, 6(2), 113-116, doi: 10.7763/IJiet.2016.V6.668
- Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practices*, 12, 15-20.
- Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21:4 361-375
- Wise, S. L., Plake, B. S, Johnson P. L. & Roos, L. L (1992). A comparison of self-adapted and computerized adaptive tests. *Journal of Educational Measurement*, 29(4), 329-339
- Zhao, R. & Wang, J. (2011). Visualizing the research on pervasive and ubiquitous computing. *Scientometrics*, 86(3),593–612