

La importancia de la analítica de datos en el seguimiento a estudiantes para el logro de certificaciones profesionales de TI: Estudio de caso

The Importance of Data Analytics in Student Tracking to Achieve IT Professional Certifications: Case Study

Carlos Alberto Baltazar Vilchis¹, Yenit Martínez Garduño², Antonio Sámano Ángeles³

Universidad Autónoma del Estado de México. Centro Universitario Atlacomulco. México
¹cabaltazarv@uaemex.mx, ²ymartinezg@uaemex.mx, ³asamanoa@uaemex.mx

Fecha de recepción: 28 de diciembre de 2020

Fecha de aceptación: 26 de abril de 2021

Resumen. La analítica de datos es una rama de la Informática que apoya a la toma de decisiones en diversas áreas del conocimiento, la educación no es la excepción, año tras año los docentes imparten sus unidades de aprendizaje en donde la aplicación de exámenes arroja un sin fin de información que, en el mejor de los casos queda almacenada en el equipo de cómputo que funja como servidor y ofrezca un Sistema para Gestión de Aprendizaje (*Learning Management System – LMS* por sus siglas en inglés) como puede ser Moodle, y queda sin utilizarse. El aplicar técnicas de análisis de datos a esta información se puede ponderar el grado de avance en cuanto al dominio de una habilidad en Tecnologías de la Información y la Computación (TIC), como pueden ser las certificaciones que otorga la empresa CISCO®, hoy sabemos que las certificaciones se han vuelto indispensables para lograr una ventaja competitiva entre los profesionistas de las TIC y dar una oportunidad para posicionarse en un buen empleo en un mundo globalizado.

El presente trabajo de investigación analiza, mediante un estudio de alcance descriptivo, comparativo y correlacional, la unidad de aprendizaje “Comunicación entre Computadoras” de las generaciones 2016 a la 2020 de los grupos de la Licenciatura en Informática Administrativa LIA D1, LIA D2, LIA D3 y LIA D4 del Centro Universitario UAEM Atlacomulco, a través de una minería de datos al sistema de reactivos de exámenes aplicados por el docente ubicados en una plataforma *LMS Moodle*, para posteriormente codificarlos y compararlos con las certificaciones CISCO® CCNA 100-101 (ICND1), 200-101 (ICND2) y 200-120 (CCNA R & S) para determinar si los discentes tienen las habilidades requeridas para presentarlas y aprobarlas.

Palabras Clave: Análisis de Datos, Informática, Certificaciones, CISCO®, CCNA.

Summary. Data analytics is a branch of Computer Science that supports decision-making in various areas of knowledge, education is no exception, year after year teachers teach their learning units where the application of exams yields endless of information that, in the best of cases, is stored in the computer equipment that acts as a server and offers a Learning Management System (LMS) such as Moodle, and remains unused. By applying data analysis techniques to this information, the degree of progress in mastering a skill in Information and Computing Technologies (ICT) can be weighted, such as the certifications granted by the CISCO® company, today we know that certifications have become essential to achieve a competitive advantage among ICT professionals and provide an opportunity to position themselves in a good job in a globalized world.

The present research work analyzes, through a descriptive, comparative and correlational study, the learning unit "Communication between Computers" from the generations 2016 to 2020 of the groups of the Bachelor's Degree in Administrative Informatics LIA D1, LIA D2, LIA D3 and LIA D4 of the UAEM Atlacomulco University Center, through data mining to the system of test items applied by the teacher located on an LMS Moodle platform, to later code and compare them with the CISCO® CCNA 100-101 certifications (ICND1), 200-101 (ICND2) and 200-120 (CCNA R & S) to determine if learners have the required skills to present and pass them.

Keywords: Data Analysis, Computing, Certifications, CISCO®, CCNA

1 Introducción

En nuestro mundo donde las organizaciones y la economía cada vez son más dinámicas, en la que las instituciones deben tomar decisiones oportunas y acertadas para llegar a sus clientes, sean alumnos, compradores, etc, y además deben seguir creciendo, el análisis de datos se ha vuelto una herramienta cada vez más importante para lograr sus objetivos.

Si bien la información que éstas obtienen –tanto de sus procesos internos y proveedores, como especialmente de sus clientes– se estudia a través de diferentes técnicas desde hace ya un tiempo, tecnologías como **Big Data** han llevado esto un paso más allá para poder utilizar datos no estructurados, pero especialmente para hacer estos análisis en tiempo real. Industrias como la financiera, la de telecomunicaciones y el retail están aprovechando este tipo de soluciones y la educación no es ajena a su implementación debido al uso de LMS como Moodle,

Schoology, EdModo, por mencionar algunos, el análisis de la información resultante del proceso evaluativo mediante estas TIC se le conoce también como Analítica del Aprendizaje (Learning Analytics, LA por sus siglas en inglés), definida como la medición, recopilación, análisis e informe de datos sobre los alumnos y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce. [1]

2 Estado del arte

La analítica del aprendizaje es un área importante de aprendizaje mejorado por la tecnología que ha surgido durante la última década en donde los factores tecnológicos, educativos y políticos han impulsado el desarrollo de la analítica en entornos educativos. La analítica basada en datos es una disciplina que surge en el siglo XX, permitiendo el surgimiento de perspectivas centradas en el aprendizaje y la influencia de las preocupaciones económicas nacionales. Luego se centra en las relaciones entre el análisis de aprendizaje, la minería de datos educativos y el análisis académico. [2]

Recientemente, existe un interés creciente en el aprendizaje mejorado por tecnología (TEL por sus siglas en inglés). En general, la analítica de aprendizaje se ocupa del desarrollo de métodos que aprovechan los conjuntos de datos educativos para apoyar el proceso de aprendizaje. LA es un campo multidisciplinario que involucra aprendizaje automático, inteligencia artificial, recuperación de información, estadísticas y visualización, también es un campo en el que convergen varias áreas de investigación relacionadas en TEL. Estos incluyen análisis académicos, análisis de acciones y minería de datos educativos. Los investigadores Mohamed Amine Chatti, Anna Lea Dyckhoff, Ulrik Schroeder y Hendrik Thüs desarrollaron una investigación sobre las conexiones entre LA y estos campos relacionados. Describieron un modelo de referencia para AL basado en cuatro dimensiones; a saber, datos y entornos (¿qué?), partes interesadas (¿quién?), objetivos (¿por qué?) y métodos (¿cómo?). Revisaron publicaciones recientes sobre LA y sus campos relacionados y las asignaron a las cuatro dimensiones del modelo de referencia. Además, identificaron varios desafíos y oportunidades de investigación en el área de AL en relación con cada dimensión. [3]

Por su parte Doug Clow desarrolló un artículo aplicando el modelo de cinco pasos de Campbell y Oblinger [4] de análisis de aprendizaje (captura, informe, predicción, actuación, refinamiento) y otras teorizaciones del campo, y se basa en una teoría educativa más amplia para articular un Ciclo de análisis de aprendizaje más desarrollado, explícito y basado en la teoría. Este ciclo conceptualiza el trabajo exitoso de análisis de aprendizaje como cuatro pasos vinculados: los alumnos (1) generan datos (2) que se utilizan para producir métricas, análisis o visualizaciones (3). El paso clave es 'cerrar el círculo' retroalimentando este producto a los alumnos a través de una o más intervenciones (4). Este documento busca comenzar a colocar la práctica de análisis de aprendizaje sobre una base de teoría de aprendizaje establecida, y extrae varias implicaciones de esta teoría para la mejora de los proyectos de análisis de aprendizaje. Estos incluyen acelerar o acortar el ciclo para que la retroalimentación ocurra más rápidamente y ampliar la audiencia para recibir retroalimentación (en particular, considerando a los estudiantes y maestros como audiencias para el análisis) para que pueda tener un mayor impacto. [5]

A nivel personal, la presente investigación agrega un nuevo grupo para su análisis, en la investigación realizada en el año 2019 se describe la construcción de un instrumento basado en las certificaciones CISCO® CCNA 100-101 (ICND1), 200-101 (ICND2) y 200-120 (CCNA R & S) [6] para analizar las bases de datos existentes en el sistema Moodle LMS de los exámenes aplicado a los estudiantes de Licenciatura en Informática Administrativa (LIA) del Centro Universitario Atlacomulco (CUA) perteneciente a la Universidad Autónoma del Estado de México (UAEM) durante los años escolares 2016 a 2019 a la asignatura "Comunicación entre computadoras" a través de una prueba previa al comienzo "*pretest*" y una prueba posterior "*posttest*" al final de cada año escolar, lo que permitirá evaluar el grado de autopercepción de competencia e identificar a los estudiantes que podrían competir por una probable certificación de esta empresa. Los resultados obtenidos proporcionaron evidencia sobre la calidad del instrumento que presentó una estructura de 11 factores. Específicamente, se obtuvo una consistencia interna entre .883 y .947 de la medición Alpha de Cronbach para cada prueba previa y posterior aplicada y el análisis factorial exploratorio, que se pretendía desarrollar, no fue necesario debido a la ipsatividad de los datos, esto motivó que el instrumento tiene evidencia de confiabilidad y validez que permite explorar las competencias de los estudiantes en esta materia. [7]

3 Metodología

La investigación retoma el modelo propuesto por Baltazar, Martínez, Sámano, Garduño y Evangelista [7] para evaluar las competencias en cuanto a las certificaciones CISCO® CCNA 100-101 (ICND1), 200-101 (ICND2) y 200-120 (CCNA R & S), un instrumento ya validado y con el nivel de confidencialidad pertinente

para la identificación de estos factores. El instrumento permitirá saber lo que saben y lo que los estudiantes son capaces de hacer. Para esto, los indicadores de las certificaciones antes mencionadas se tomaron como referencia a través de una escala de calificación tipo Likert, para valorar los resultados obtenidos de las pruebas realizadas en una escala de 1 a 10 su nivel de conocimiento y habilidades, donde el valor 1 se refiere a que el estudiante es completamente ineficaz en hacer lo que se presenta, y 10 el dominio completo de la declaración.

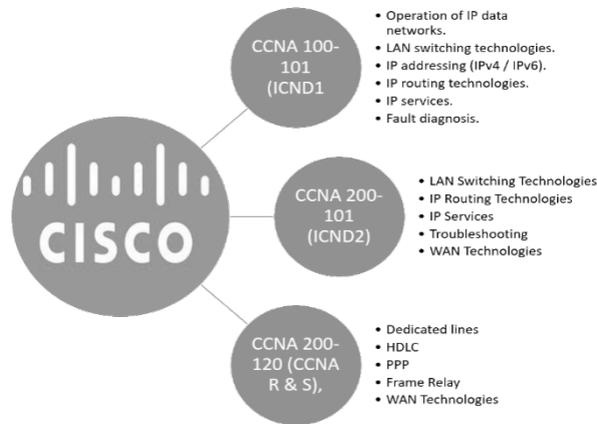


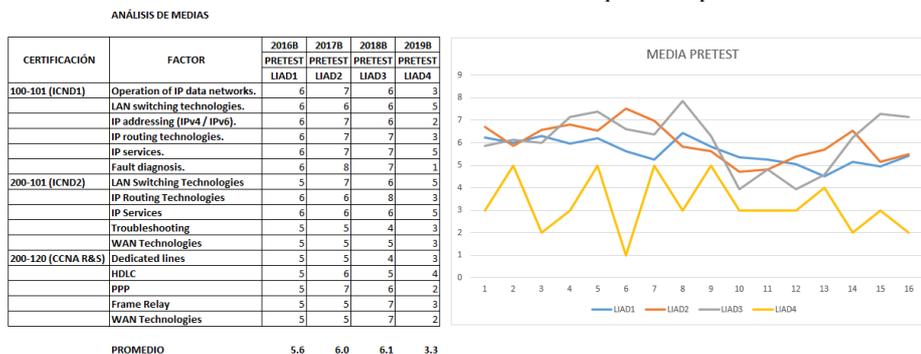
Fig. 1. Variables de las Certificaciones CISCO®

Las bases de datos del sistema Moodle del UAEM se extrajeron de las evaluaciones realizadas al tema "Comunicación entre computadoras", donde cada pregunta se colocó dentro de los dieciséis factores del instrumento, arrojando un total de 80 reactivos que se analizaron utilizando el programa SPSS 25, se definirán los promedios de cada variable y su dispersión, más adelante el coeficiente alfa de Cronbach determinará su consistencia interna, el índice de adecuación de la muestra de Kaiser-Meyer-Olkin (KMO) [8], para comprobar si la adecuación de los datos es relevante para realizar el Análisis Factorial Exploratorio (AFE) [9] y la prueba de esfericidad de Bartlett [10] que indicará el buen ajuste del modelo. Finalmente, su solución factorial se obtendrá a través de la rotación Varimax. [11], el instrumento se aplicó en dos etapas, a través de una prueba previa al comienzo de cada curso y una prueba posterior al final, se determinaron medidas de tendencia central, dispersión y análisis correlacional para obtener posteriormente el AFE.

4 Resultados

La Tabla 1 muestra en análisis de la dispersión de los promedios de las dimensiones del modelo aplicado para la etapa "pretest", se distingue de manera clara el desempeño de cada generación que cursó la materia evaluada, por poco el grupo de LIAD3 con un promedio de 6.1 se pondera como el mejor al iniciar sus estudios, el grupo que menores habilidades denotó es el reciente LIAD4 con un muy bajo promedio de 3.3, una de las posibles causas de este resultado es el cambio de docente a éste grupo para el ciclo escolar 2019B quien con otro enfoque, muy propio de cada docente, impartió dicha unidad de aprendizaje.

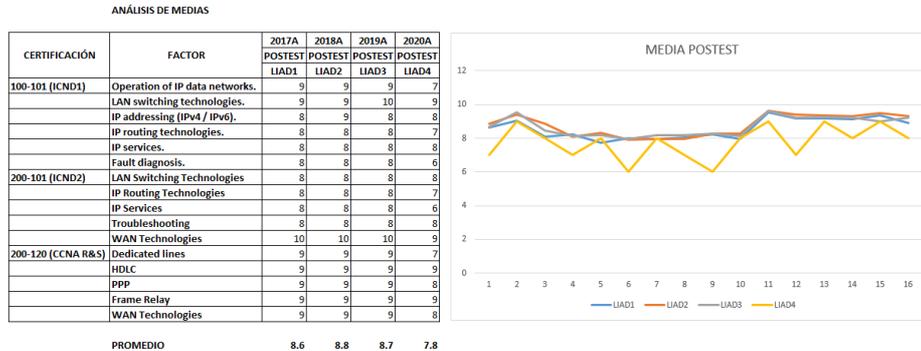
Tabla 1. Resultados de la aplicación pretest



Fuente: Elaboración propia

La Tabla 2 muestra la dispersión de los resultados del análisis de la media del “postest”, donde de manera clara se denota la mejora en los grupos al fin de su ciclo, los grupos LIA D1, D2 y D3 por pocas décimas quedaron de diferencia uno entre otro quedando en primer lugar el grupo D2 como mejor posicionado, el grupo LIA D4 mejoró un 69%, sin embargo hasta el momento es el que menor resultado ha mostrado de habilidades para una posible certificación profesional. También es posible identificar la ipsatividad de los datos entre las primeras tres generaciones, la última muestra un poco mayor dispersión.

Tabla 2. Resultados de la aplicación postest



Fuente: Elaboración propia

Con base en los datos de la Tabla 3, se valida una mejora en los resultados durante cada año escolar hasta el ciclo 2019B - 2020A, el grupo LIA D1 tenía una inscripción de 26 estudiantes, solo 2 tendrían la posibilidad de obtener una certificación CISCO®, 15 estudiantes podrían alcanzarla si su rendimiento mejorara y 9 serían descartados, el grupo LIA D2 tenía 40 estudiantes de los cuales 7 podrían lograrla mientras que 22 estudiantes estarían en riesgo y 11 serían descartados.

El grupo LIA D3, 6 estudiantes pudieron obtener una certificación CISCO, mientras que 11 estudiantes estarían en riesgo y 8 serían descartados. Se identifica que se lograrían certificar a 6 estudiantes, un porcentaje es más alto que el resto con un 24%, disminución de los estudiantes en riesgo con un 44% y estudiantes descartados con un 8%.

Finalmente LIA D4 sería una generación en peligro de obtener la certificación CISCO, de 22 estudiantes matriculados sólo 1 (4.5%) estaría en posibilidades de presentar los exámenes de habilidades, 9 alumnos en riesgo (40.9%) y 12 descartados (54.5%) más de la mitad del grupo.

Tabla 3. Análisis de posibles estudiantes certificados por año escolar.

GRUPO	CICLO ESCOLAR	TOTAL ESTUDIANTES	POSIBLE CERTIFICACIÓN	%	EN RIESGO	%	DESCARTADOS	%
LIA D1	2016B - 2017A	26	2	7.7%	15	57.7%	9	34.6%
LIA D2	2017B - 2018A	40	7	17.5%	22	55.0%	11	27.5%
LIA D3	2018B - 2019A	25	6	24.0%	11	44.0%	8	32.0%
LIA D4	2019B - 2020A	22	1	4.5%	9	40.9%	12	54.5%

Fuente: Elaboración propia

La Tabla 4 muestra el coeficiente Alfa de Cronbach de cada periodo evaluado, el promedio que se obtuvo en la presente investigación fue de .887 indicando una consistencia interna total satisfactoria así una correlación muy alta la cual no mejoraría si se eliminase algún ítem del instrumento. Por tanto, se puede afirmar que los índices de consistencia interna obtenidos en las distintas puntuaciones son aceptables, tanto para el conjunto de ítems como para cada bloque de ítems.

Tabla 4. Alfa de Cronbach por grupo y ciclo escolar

CERTIFICACIÓN	FACTOR	LIA D1		LIA D2		LIA D3		LIA D4	
		2016B	2017A	2017B	2018A	2018B	2019A	2019B	2020A
		POSTEST							
100-101 (ICND1)	Operation of IP data networks.	0.861	0.905	0.927	0.828	0.775	0.863	0.919	0.839
	LAN switching technologies.	0.872	0.827	0.816	0.879	0.905	0.933	0.927	0.919
	IP addressing (IPv4 / IPv6).	0.860	0.896	0.843	0.867	0.846	0.852	0.886	0.927
	IP routing technologies.	0.850	0.839	0.918	0.853	0.899	0.870	0.913	0.886
	IP services.	0.761	0.919	0.852	0.856	0.785	0.866	0.919	0.913
200-101 (ICND2)	Fault diagnosis.	0.880	0.927	0.944	0.916	0.929	0.917	0.949	0.919
	LAN Switching Technologies	0.807	0.886	0.920	0.820	0.783	0.829	0.938	0.944
	IP Routing Technologies	0.773	0.913	0.811	0.853	0.932	0.806	0.920	0.920
	IP Services	0.794	0.919	0.805	0.856	0.785	0.866	0.811	0.811
	Troubleshooting	0.804	0.949	0.912	0.849	0.782	0.864	0.805	0.805
200-120 (CCNA R&S)	WAN Technologies	0.763	0.938	0.830	0.843	0.757	0.828	0.912	0.912
	Dedicated lines	0.881	0.816	0.865	0.804	0.785	0.817	0.820	0.932
	HDLC	0.861	0.951	0.823	0.832	0.855	0.810	0.853	0.785
	PPP	0.926	0.876	0.850	0.875	0.922	0.875	0.856	0.782
	Frame Relay	0.940	0.938	0.940	0.783	0.913	0.892	0.849	0.757
	WAN Technologies	0.916	0.926	0.855	0.852	0.761	0.832	0.843	0.899
PROMEDIO ALFA DE CRONBACH		0.847	0.902	0.869	0.848	0.838	0.858	0.883	0.872

Fuente: Elaboración propia

Al tratar de obtener la matriz de correlaciones obtenida del programa SPSS de cada grupo el programa informático indicó que "esta matriz no es definida positiva", por lo que nuevamente se determinó que no era necesario realizar el análisis factorial, ya que la escala de los resultados utilizados son ipsativos y los resultados están interrelacionados, por lo tanto, se afirma que, dado que la matriz no es positiva, se supone que el instrumento utilizado es válido. Cattell [12] inició el uso del término "ipsativo"; (del latín ipse: él mismo) para referirse a las transformaciones de puntaje sin procesar que centran los puntajes sobre la media del individuo. McLean, JE, Chissom, BS [13] comentan que las escalas Ipsativas han sido empleadas principalmente por investigadores en el área de la medición de la personalidad, la elección vocacional y la evaluación de valores y actitudes donde los valores de la escala están interrelacionados y el presente estudio está dentro de estos .

5 Conclusiones

Desde los inicios de la Informática la gestión de la información siempre ha sido importante debido a que ésta puede ofrecerle al tomador de decisiones la oportunidad de lograr una ventaja competitiva independientemente del ámbito en donde se aplique, el análisis de los datos que se desarrolló en la presente investigación permite llegar a las siguientes conclusiones; 1) Analizar los datos de los discentes permitirá al docente identificar sus áreas de oportunidad en beneficio de sus educandos. 2) Es importante basar siempre y en la medida de lo posible, la temática académica vista en el aula contra una certificación, en el caso de esta investigación es posible debido a que el área de las TIC, al ser tan dinámica, necesita profesionales con las habilidades suficientes para su administración, ésta deberá cambiar efectivamente dependiendo el área y las posibles certificaciones que tenga. 3) El apearse a un plan de estudio no siempre significa que el alumno esté desarrollando habilidades necesarias para su vida laboral, en el caso del grupo LIA D4, al cambiar el docente cambia la forma de enseñanza y si no se centra el esfuerzo académico por parte del profesor en acercar temas de vanguardia, con mucho mayor razón alguna certificación internacional, en vez de ayudar al alumno se le está perjudicando. 4) Es importante mencionar que al grupo LIA D4 se dejó de atender de manera presencial en el mes de marzo de 2020 y las actividades cambiaron a virtuales debido a la contingencia sanitaria del COVID-19 que se suscitó en México y en el mundo, y debieron volverse en cierta forma autodidactas, porque a pesar de tener los LMS o sistemas para videoconferencias es innegable que si el alumno no aprende por su propia cuenta, si no va más allá del camino que su docente le pueda indicar, poco sirve la tecnología y la teoría educativa más vanguardista. 5) Finalmente en cuanto al instrumento de medición que se aplicó en el presente trabajo debe contemplar otras variables que permitan una analítica de datos más robusta y permita identificar a mayor profundidad el escenario idóneo para que un alumno logre los aprendizajes esperados, como físicas, sociales y económicas. Se debe desarrollar una investigación con mayor amplitud en beneficios de los discentes.

Referencias

- [1]. Universidad Virtual de Valencia. La analítica en el entorno educativo: ventajas para su implementación. Recuperado de <https://www.universidadviu.com/la-analitica-entorno-educativo-ventajas-implementacion/>. 2018

- [2]. R. Ferguson. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*. Print ISSN: 1753-5255 Online ISSN: 1753-5263. 2013
- [3]. M. A. Chatti, A. L. Dyckhoff, U. Schroeder y Hendrik Thüs. A reference model for learning analytics. *International Journal of Technology Enhanced Learning*. Print ISSN: 1753-5255 Online ISSN: 1753-5263. 2013
- [4]. Campbell, J. P. and Oblinger, D. G. Academic Analytics. *EDUCAUSE Quarterly*. October (2007). 2007.
- [5]. D. Clow. The learning analytics cycle: closing the loop effectively. *LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* April 2012 Pages 134–138 <https://doi.org/10.1145/2330601.2330636>. 2012
- [6]. E. Arguello. *Redes Cisco. Guía de estudio para la certificación CCNA Routing y Switching*. Editorial RA-MA. ISBN. 978-84-9964-272-7
- [7]. C.A.Baltazar, Y. Martínez, A. Sámano, A. Garduño, F. G. Corte, E. Evangelista. Proposal of an instrument for measuring educational quality based on the CISCO® CCNA 100-101 (ICND1), 200-101 (ICND2) and 200-120 (CCNA R&S) certifications. In *ACM International Conference Proceeding Series (Vol. 101, pp. 205–211)*. <https://doi.org/10.1145/3369199.3369202>. 2019
- [8]. Hair, J, Black, W, Babin, B. *Multivariate data analysis*, 7th Ed. Upper Saddle River, NJ: Prentice-Hall. 2009
- [9]. Pett, M, Lackey, N, Sullivan, J. *Making sense of factor analysis*. Thousand Oaks, CA: Sage. 2003
- [10]. Bartlett, M. Tests of significance in factor analysis. *Br J Psychol*. 1950
- [11]. Gorsuch, R. *Factor analysis* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 1983
- [12]. Cattell, R. B. 1952 Factor Analysis. New York: Harper & Row. "Psychological measurement: ipsative, normative and interactive." *Psych. Rev. Rev.* 51: 292- 303.
- [13]. McLean, J. E., Chissom, B. S. 1986. *Multivariate Analysis of Ipsative Data: Problems and Solutions*. Annual Meeting of the Mid-South Educational Research Association Memphis, TN, November 19-21, 1986, pp. 83-94.