

# 23.EVALUACIÓN DE LOS APRENDIZAJES CON INTELIGENCIA ARTIFICIAL ASSESSMENT OF LEARNING WITH ARTIFICIAL INTELLIGENCE

*Alberto Valdez Sandoval<sup>39</sup>*

***Fecha recibida:*** 27/09/2022

***Fecha aprobada:*** 15/12/2022

***Derivado del proyecto:*** Sistema Experto de Evaluación Inteligente SEEI

***Pares evaluadores:*** Red de Investigación en Educación, Empresa y Sociedad – REDIEES.

---

<sup>39</sup> Ingeniero en Computación, Universidad de Guadalajara, Doctorado en Educación Basada en Competencias, Centro Escolar Mar de Cortés, Ocupación (Docente-Investigador), Centro de Bachillerato Tecnológico industrial y de servicios CBTis 224, correo electrónico: [alberto.valdez@cbtis224.edu.mx](mailto:alberto.valdez@cbtis224.edu.mx).

## RESUMEN

En el presente trabajo se comparan resultados obtenidos del *software* de sistema experto publicado por Sánchez et al. (2021) contra los resultados arrojados por el Sistema Experto de Evaluación Inteligente (SEEI). Este último es la actualización del primero, por lo que en este documento se exponen las similitudes y diferencias entre el *software* predecesor y el sucesor.

Ambos *softwares* utilizan Inteligencia Artificial (IA) para calcular en una escala de cero a diez, la calificación resultante de un proceso de evaluación de los aprendizajes de estudiantes de bachillerato, en el contexto mexicano.

El objetivo de esta investigación es encontrar validez en el instrumento SEEI, ya que este evolucionó de su predecesor con intenciones de que pueda ser utilizado por cualquier docente, para evaluar cualquier competencia, en cualquier asignatura, en cualquier institución educativa situada en cualquier parte del mundo con sólo tener acceso a internet.

Los resultados dan evidencia firme que la línea de investigación va por buen camino y que merece la pena continuar realizando pruebas y ajustes a SEEI y encontrar en un futuro, validez en más contextos.

De esta manera se reafirmaría el poder de cómputo de la IA sobre los modelos matemáticos basados en rúbrica con los que actualmente se evalúan los aprendizajes basados en competencia y surgiría una nueva y prometedora era en las que los docentes por fin tendrán herramientas de evaluación de vanguardia.

**PALABRAS CLAVE:** *Lógica Difusa, Sistema Experto, Evaluación, Competencia, Rúbrica.*

## ABSTRACT

In this article are compared the obtained results of the expert system *software* published by Sánchez et al. (2021) against the obtained results acquired by Sistema Experto de Evaluación Inteligente (SEEI). The last one is in fact, an update of the first one. The present document exposes the similarities and differences between predecessor and successor *software*.

Both *software* uses Artificial Intelligence (AI) to calculate, in zero to ten scale, the obtained grade of an evaluation learning process for high school students in a mexican context.

The goal in this research is to find validity in the SEEI instrument, because this one evolved from his predecessor so any teacher will be able to use it in order to evaluate any competency on any subject, in any school and around the world using simply internet connection.

The results found show solid evidence about the importance to continue in this line of research with the intention of getting better adjusts to SEEI and find in a future, validity in different contexts.

If this happens, the power of the AI algorithms will be used to replace the obsolete math models based on rubrics that are actually used for the learning assessments with the approach of competencies and will emerge a new era in which the teachers will have access to new class tools for evaluating their students.

**KEYWORDS:** *Fuzzy Logic, Expert System, Assessment, Competency, Rubric.*

## INTRODUCCIÓN

De las ciencias computacionales se desprenden varias ramas, entre ellas, la Inteligencia Artificial (IA) y de ella también se ramifican algunas otras áreas de estudio, como es la Lógica Difusa (LD). Esta última permite resolver problemas computacionales en los que existe incertidumbre, ya que los modelos matemáticos que esta disciplina aporta se utilizan con la finalidad de que la computadora pueda tomar decisiones de manera aproximada a como lo hace el cerebro humano. La LD es ideal para resolver problemas en los que se implique subjetividad.

Gutiérrez y Ferreira (2020) definen la LD como un acercamiento entre la precisión de las matemáticas clásicas y la imprecisión del mundo real, y afirman que, si esta imprecisión se intenta llevar a otro modelo matemático que no la permite (la imprecisión), entonces se obtienen resultados indeseables (p.2).

Por su parte Jamsandekar y Mudholkar (2013) reafirman la idea al mencionar que una de las formas más apropiadas para manejar múltiples variables imprecisas es utilizando LD para reflejar el modo de reflexionar de un humano, incluso en medio de incertidumbre (p.160).

En la actualidad los docentes no cuentan con instrumentos de evaluación de aprendizajes de vanguardia, por el contrario, se emplean obsoletas rúbricas o algún otro instrumento basado en escalas.

Las rúbricas son básicamente matrices de doble entrada fáciles de diseñar y aplicar, en ellas se busca la integración de la evaluación cuantitativa mediante escalas, con la evaluación cuantitativa al establecer y definir explícitamente criterios de los rubros (Cano, 2015; Gatica y Uribarren, 2012; Martínez, 2008; Picón, 2013).

Aún con la buena intención de estos instrumentos, en el mayor de los casos no entregan una buena aproximación, por ejemplo, suceden situaciones variadas como las siguientes:

- a) Alumna A y Alumna B obtuvieron la misma calificación al aplicar la rúbrica, sin embargo, el docente considera que Alumna B tuvo un mejor desempeño y el instrumento no lo reflejó.

- b) Alumno C obtuvo con el instrumento un 6 (calificación mínima aprobatoria), sin embargo, el docente considera que el estudiante todavía no alcanza el mínimo necesario para el desarrollo de la competencia.

Estas aproximaciones deficientes se deben a que los instrumentos aplicados utilizan modelos matemáticos que no se asemejan a los cálculos mentales que suceden dentro del cerebro del docente cuando este emite un juicio de valor.

En el estudio realizado por Sánchez et al. (2021) se muestran los hallazgos al utilizar un *software* que aplica LD para evaluar aprendizajes por competencias en la asignatura de probabilidad y estadística. El documento revela que el *software* incrementó significativamente la precisión y exactitud comparado con una rúbrica que emplea promedio y otra que utiliza regla de tres.

El propósito de la presente investigación es comparar los resultados obtenidos del mencionado estudio, contra los resultados que arroja la actualización de ese *software*, ahora llamado Sistema Experto de Evaluación Inteligente (SEEI), para determinar si estos ajustes al modelo difuso siguen teniendo validez en términos de precisión matemática en SEEI.

La importancia de realizar la comparación radica en que ahora SEEI fue modificado, no sólo con ajustes al modelo para evaluar el desempeño en probabilidad y estadística, también fue dotado de un algoritmo que adapta el cerebro no biológico de LD a cualquier asignatura, en cualquier nivel educativo y para evaluar cualquier competencia. Así, los resultados favorables de este estudio muestran evidencia suficiente para que se continúe con la línea de investigación y empezar a validar SEEI en esos otros contextos.

## MATERIAL Y MÉTODOS

La presente investigación se circunscribe en el enfoque cuantitativo, que este a su vez se inspira en el positivismo, el cual es una corriente filosófica que según Monje (2011) “Lo que importa para el positivismo es la cuantificación, la medición” (p. 12).

En este enfoque se utiliza el método hipotético deductivo, que va de lo general a lo particular, el punto de partida son las leyes universales o teorías causales que expliquen la realidad que experimenta el investigador. Los estudios cuantitativos se fundamentan en estudios previos, por lo que la literatura es importante para definir teorías, hipótesis y el

diseño mismo de la investigación. (Del Canto y Silva, 2013; Hernández, Fernández y Baptista, 2010).

El diseño de investigación aplicado, el plan a seguir para lograr los objetivos, responder a las preguntas de investigación y analizar la certeza de las hipótesis (Hernández, Fernández y Baptista, 2010) es de tipo no experimental, ya que no se pretendió tener control sobre las variables de estudio ni se tuvieron grupos de control.

Respecto al muestreo y debido a que la presente investigación se enfoca en los resultados que arrojó el *software* predecesor contra los resultados que arrojó SEEI, se extrajeron de la base de datos del *software* predecesor todas las calificaciones de entrada al sistema difuso, de los 143 estudiantes inscritos en la asignatura de probabilidad y estadística impartida en el sexto semestre del ciclo escolar Febrero – Julio 2017 en el Centro de Estudios Tecnológicos industrial y de servicios CETis 108. Una vez extraída la base de datos se procedió a alimentar SEEI con la misma información, para así poder realizar la comparación de los resultados.

La competencia en cuestión es: Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación. Esta se dividió conceptualmente en tres dimensiones: (a) teoría, (b) práctica, (c) actitud. Y a su vez en seis atributos de competencia mostrados en la Tabla 1.

**Tabla 1. Atributos que se evaluaron en los estudiantes para la competencia**

ATRIBUTO	DIMENSIÓN
1. Lleva los conocimientos teóricos a la práctica.	Teoría.
2. Organiza, clasifica, presenta e interpreta información numérica en forma de tablas y gráficas.	Práctica.
3. Calcula e interpreta índices estadísticos como medidas de tendencia central y medidas de dispersión.	Práctica.
4. Presenta los resultados de los índices estadísticos con honestidad, pulcritud de los datos y exactitud.	Actitud.
5. Utiliza las TIC como medio para gestionar su aprendizaje.	Práctica.
6. Cumple con sus actividades con honestidad y puntualidad.	Actitud.

Ambos instrumentos de medición fueron elaborados con el modelo matemático basado en LD tipo Mandami. Para el modelado se utilizó la herramienta Fuzzy Logic Toolbox de Matlab R2013a. Se modelaron las funciones de membresía de las entradas/atributos y de la salida calificación. Las funciones para el *software* predecesor se detallan en la Tabla 2 y Tabla 3 respectivamente.

**Tabla 2. Funciones de membresía de entrada del modelo difuso predecesor**

ENTRADAS/ATRIBUTOS	FUNCIONES DE MEMBRESÍA
1. Lleva los conocimientos teóricos a la práctica. Teoría.	Bajo: $y = trimf(x, [0\ 0\ 4])$ Medio: $y = trapmf(x, [3\ 5\ 6\ 6.5])$ Alto: $y = gaussmf(x, [0.927\ 8.85])$
2. Organiza, clasifica, presenta e interpreta información numérica en forma de tablas y gráficas. Práctica.	Bajo: $y = trimf(x, [0\ 0\ 3])$
4. Presenta los resultados de los índices estadísticos con honestidad, pulcritud de los datos y exactitud. Actitud.	Regular: $y = trimf(x, [1\ 3\ 5])$ Bueno: $y = gbellmf(x, [1.05\ 1.656\ 6.169])$
6. Cumple con sus actividades con honestidad y puntualidad. Actitud.	Alto: $y = gaussmf(x, [1\ 9])$
3. Calcula e interpreta índices estadísticos como medidas de tendencia central y medidas de dispersión. Práctica.	Bajo: $y = trimf(x, [0\ 0\ 3])$ Medio: $y = trapmf(x, [2\ 5\ 6\ 8])$
5. Utiliza las TIC como medio para gestionar su aprendizaje. Práctica.	Alto: $y = gaussmf(x, [1.699\ 10])$

**Tabla 3. Funciones de membresía de salida del modelo difuso predecesor**

SALIDA	FUNCIONES DE MEMBRESÍA
Calificación	Bajo: $y = gaussmf(x, [1.293\ 0])$ Regular: $y = gaussmf(x, [0.9373\ 3.72])$ Bueno: $y = gaussmf(x, [0.7958\ 7.66])$ Alto: $y = gaussmf(x, [0.9772\ 9])$

A fin de mostrar los ajustes al modelo predecesor, con la que se logró la actualización llamada SEEI, la Tabla 4 describe cómo se reorganizaron las funciones de membresía para las entradas. En la variable de salida calificación no se hicieron modificaciones.

**Tabla 4. Funciones de membresía de entrada del modelo difuso de SEEI**

ENTRADAS/ATRIBUTOS	FUNCIONES DE MEMBRESÍA
1. Lleva los conocimientos teóricos a la práctica. Teoría.	Bajo: $y = trimf(x, [0\ 0\ 4])$ Medio: $y = trapmf(x, [3\ 5\ 6.5\ 7])$ Alto: $y = gaussmf(x, [1.5\ 10])$
2. Organiza, clasifica, presenta e interpreta información numérica en forma de tablas y gráficas. Práctica.	Bajo: $y = trimf(x, [0\ 0\ 3])$
3. Calcula e interpreta índices estadísticos como medidas de tendencia central y medidas de dispersión. Práctica.	Medio: $y = trapmf(x, [2\ 5\ 6\ 8])$ Alto: $y = gaussmf(x, [1.699\ 10])$
5. Utiliza las TIC como medio para gestionar su aprendizaje. Práctica.	
4. Presenta los resultados de los índices estadísticos con honestidad, pulcritud de los datos y exactitud. Actitud.	Bajo: $y = trimf(x, [0\ 0\ 3])$ Regular: $y = trimf(x, [1.1672\ 3.672\ 5.672])$ Bueno: $y = gbellmf(x, [1.05\ 1.66\ 6.678])$
6. Cumple con sus actividades con honestidad y puntualidad. Actitud.	Alto: $y = gaussmf(x, [1\ 9.964])$

Además de algunos ajustes evidentes en los argumentos de las funciones de pertenencia, que se pueden observar entre la Tabla 2 y Tabla 4, también se reorganizaron de tal manera que todos los atributos agrupados a cada dimensión (teoría, práctica o actitud) tuvieran las mismas funciones de pertenencia, esto fue para dotar a SEEI con un algoritmo

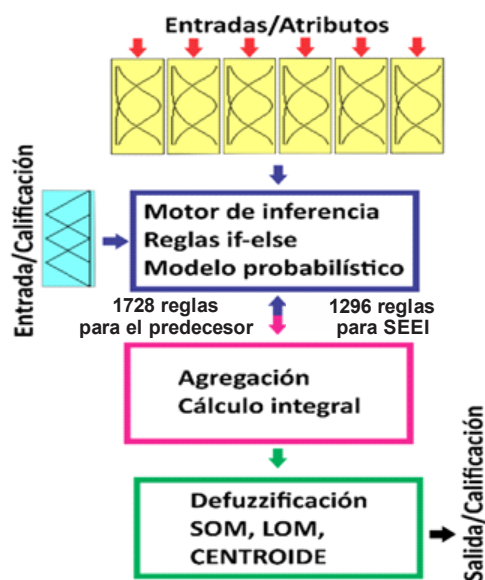
que es capaz de reestructurarse o adaptarse para cada competencia que cualquier docente desee evaluar, sin embargo, esta característica quedó fuera de los alcances de este estudio.

Las reglas de inferencia para los dos algoritmos fueron generadas mediante un modelo probabilístico de combinaciones. Para el sistema predecesor, por regla de multiplicación se obtuvieron un total de 1728 reglas al multiplicar  $3 \times 4 \times 3 \times 4 \times 3 \times 4$ . Esto debido a que los Atributos 1, 3 y 5 tienen 3 niveles (Bajo, Medio y Alto); los Atributos 2, 4 y 6 tienen 4 niveles (Bajo, Regular, Bueno y Alto).

Siguiendo el mismo principio antes mencionado se generaron para SEEI  $3 \times 3 \times 3 \times 4 \times 3 \times 4$  ya que casi todos los atributos tienen 3 niveles de desempeño y sólo los atributos 4 y 6 tienen 4 niveles de desempeño (Tabla 4). El resultado fue un total de 1296 reglas para SEEI.

En ambos modelos difusos se utilizaron los métodos de defuzzificación SOM (*Smallest Of Maximum*), LOM (*Largest Of Maximum*) y centroide. En la Figura 1 se pueden observar el diagrama a bloques de los sistemas.

**Figura 1. Estructura de ambos modelos matemáticos de LD**



Una vez que se obtuvieron los resultados arrojados por SEEI, se procedió a someter a prueba las siguientes hipótesis para el análisis de similitudes y diferencias.

Hipótesis de diferencia entre grupos:

$H_i$ . Ambos modelos difusos conservan una precisión similar, con una diferencia menor a 5 unidades porcentuales.

Hipótesis estadística: ANOVA:

$H_0$ : ( $\mu_1 = \mu_2$ ). No hay diferencia en las medias de las mediciones calculadas con los dos instrumentos.

$H_1$ : ( $\mu_1 \neq \mu_2$ ). Existe diferencia en las medias de las mediciones calculadas con los dos instrumentos.

Se aplicaron en el diseño metodológico estadística descriptiva y prueba de hipótesis ANOVA con test de Tukey para dar respuesta a las hipótesis planteadas y así comprender las diferencias y similitudes de los dos modelos difusos.

## RESULTADOS

Los resultados arrojados por ambos *softwares* se muestran en la Tabla 5, específicamente en las columnas DIFUSO PREDECESOR y SEEI. Además, en la columna DIFERENCIA se observa que sólo 13 de los 143 resultados tuvieron un cambio significativo de una unidad o más (marcados con gris), lo que equivale al 9.09% de los resultados.

**Tabla 5. Calificaciones resultantes entre ambos modelos difusos**

ALUMNO	ATRIB.1	ATRIB.2	ATRIB.3	ATRIB.4	ATRIB.5	ATRIB.6	DIFUSO PREDECESOR	SEEI	DIFERENCIA
1	6	7	5	7	5	5	7.2	6.3	0.9
2	9	10	10	9	10	9	10	10	0
3	9	10	9	10	9	10	10	10	0
4	9	9	9	10	8	10	10	10	0
5	7	8	7	7	6	5	7.2	7.7	-0.5
6	6	7	7	6	5	6	7.8	7.7	0.1
7	9	9	8	9	9	10	10	10	0
8	8	9	8	10	10	10	10	10	0
9	7	7	7	7	7	6	7.7	7.7	0
10	4	5	6	6	7	6	7.2	6.3	0.9
11	5	6	7	6	8	7	7.8	7.7	0.1
12	3	4	5	6	6	5	5.1	5.1	0
13	4	5	6	6	7	6	7.2	6.3	0.9
14	6	7	7	7	6	7	7.8	7.7	0.1

15	7	8	8	7	8	8	7.9	7.7	0.2
16	8	9	10	10	10	10	10	10	0
17	6	8	7	7	7	7	7.8	7.7	0.1
18	9	9	9	10	9	10	10	10	0
19	9	9	10	10	9	10	10	10	0
20	4	5	4	3	2	4	4	4.1	-0.1
21	9	9	10	8	9	9	10	10	0
22	4	5	7	6	4	6	7.2	6.3	0.9
23	9	9	10	10	10	10	10	10	0
24	7	7	9	8	9	8	7.9	8.5	-0.6
25	5	4	5	7	4	6	4.3	6.3	-2
26	7	6	8	8	8	7	7.7	7.7	0
27	10	10	10	10	10	10	10	10	0
28	8	9	9	10	9	10	10	10	0
29	8	9	8	9	8	9	8.7	9	-0.3
30	4	5	4	4	5	5	7.2	5.3	1.9
31	7	7	6	6	7	7	7.7	7.7	0
32	3	4	3	3	3	3	3.8	4	-0.2
33	4	4	3	4	6	5	5.1	5.2	-0.1
34	6	5	6	6	4	6	7.2	6.3	0.9
35	5	6	6	5	5	6	7.2	6.3	0.9
36	4	5	6	6	5	6	7.2	6.3	0.9
37	10	10	10	10	10	10	10	10	0
38	7.5	7	6	7.5	7	8	7.8	7.6	0.2
39	0	0	0	0	0	0	0	0	0
40	9	9	9	8	8.5	9	8.8	9	-0.2
41	9.5	9	9.5	9	9	9.5	9.7	10	-0.3
42	7.5	8	8	7	7.5	8	7.8	7.6	0.2
43	7	7.5	8	7	7.6	7	7.9	7.6	0.3
44	3	5	2	3	2	4	4	4	0
45	8	7.5	8	7	8	7.5	7.8	7.6	0.2
46	8	8.5	8	7.9	8	8.5	8.7	8.9	-0.2
47	5	7	6	5	4	3	7.2	5.3	1.9
48	7.7	8	8.5	7	8	8	7.8	7.6	0.2
49	9	8	8.5	7.9	9	9	8.7	9	-0.3
50	8	9	9	8	9	9	8.8	9	-0.2
51	8	8.5	9	8	8	9	8.7	9	-0.3
52	10	10	10	10	9.5	9.5	10	10	0
53	8	7.7	8.9	9.3	9	8	8.7	8.9	-0.2
54	9	9	9	9.5	9	9	9.6	10	-0.4
55	6	5	7	3	3	4	5.1	5.2	-0.1
56	6	5	6	7	5	5	7.2	6.3	0.9
57	6	6	7	7.6	8	5	7.2	7.7	-0.5
58	8	8	9	9	9	9	8.8	9	-0.2

59	5	4	5	4	6	5	5.1	5.3	-0.2
60	6	7	7.5	8	7	6	7.8	7.6	0.2
61	8	8	8	7	9	7	7.8	7.7	0.1
62	8	9	8	9	7	7	7.8	8.5	-0.7
63	7	8	8	9	7	6	7.7	7.7	0
64	8	8	9	7.8	8	8	8.7	8.5	0.2
65	7	7.5	8	8	8	8	8.4	7.6	0.8
66	8	7	8	7	8	7	7.8	7.7	0.1
67	8	8	8	7	8.5	7	7.8	7.7	0.1
68	8	8	7.5	7	5	8	7.8	7.6	0.2
69	9	9	9	8.5	9	9	8.9	8.9	0
70	5	7	7	5	6.8	5	7.2	5.3	1.9
71	7	6	7.5	7	7.5	7	7.7	7.6	0.1
72	9	9	9.5	10	9	9	10	10	0
73	6	7	7	6.5	7	6	7.8	7.7	0.1
74	5	6	5	6	7	7	7.8	7.7	0.1
75	8	8	9	8	8	9	8.7	9	-0.3
76	6	5	6	5	7	7	7.2	7.7	-0.5
77	9	9	10	9	9	10	10	10	0
78	10	10	10	10	10	9.5	10	10	0
79	9	9	8	8	9	7	7.8	8.5	-0.7
80	9	8	8	9	9	9	8.7	9	-0.3
81	10	9	10	10	10	9.5	10	10	0
82	8	9	7	8	8	9	7.9	8.5	-0.6
83	5	6	4	5	6	4	7.2	5.3	1.9
84	6	7	6	7	7	7	7.8	7.7	0.1
85	7	6	7	5	6	6	7.2	7.7	-0.5
86	6	7	7	6	6	6	7.8	7.7	0.1
87	2	2	2	1	0	2	1.1	1	0.1
88	7	7	8	6	7	8	7.7	7.7	0
89	9	9	9	9	10	9	9.6	10	-0.4
90	3	2	1	1	0	1	1.1	1	0.1
91	10	9	10	8	8	9	10	10	0
92	10	9	9	9	10	10	10	10	0
93	3	2	1	1	2	1	1	1	0
94	5	4	6	4	6	4	3.8	5.4	-1.6
95	5	3	4	4	3	4	3.8	4.1	-0.3
96	7	6.5	6	7.5	7	6.5	7.8	7.7	0.1
97	5	5.7	4.5	5	5.9	5	7.2	5.3	1.9
98	6	5	6	6	5	6	7.2	6.3	0.9
99	9	8	9	8	8	6	7.8	8.5	-0.7
100	6	6	7	7	7	7	7.8	7.7	0.1
101	7	6	6	7	7	7	7.7	7.7	0
102	8	7	8	9	8	9	7.8	9	-1.2

103	6	6	7	7	6	6	7.8	7.7	0.1
104	8	9	8	8	8	7	7.8	8.5	-0.7
105	3	3	4	3	3	3	3.8	4	-0.2
106	5	4	5	3	3	5	5.1	5.2	-0.1
107	5	4	4	4	6	4	3.8	5.4	-1.6
108	4	4	5	4	4	4	3.8	5.3	-1.5
109	6	7	6	7	7	7	7.8	7.7	0.1
110	4	5	3	4	3	3	4.6	4.1	0.5
111	8	7	7	8	9	9	7.8	8.5	-0.7
112	4	5	4	4	4	3	5.1	5.3	-0.2
113	6	7	8	8	7	8	7.8	7.7	0.1
114	4	5	5	5	5	5	7.2	5.3	1.9
115	10	10	10	10	10	10	10	10	0
116	8	8	8	10	9	10	8.7	9	-0.3
117	8	7	7	8	7	8	7.8	7.7	0.1
118	10	10	10	10	10	10	10	10	0
119	9	9	10	10	10	10	10	10	0
120	3	4	3	3	3	5	4.6	4	0.6
121	6	6	6	6	7	6	7.8	7.7	0.1
122	8	9	9	10	10	10	10	10	0
123	5	4	6	6.5	6	5.7	4	6.3	-2.3
124	7	6.8	7	6	7	7.9	7.7	7.7	0
125	9	8	8.9	9	8.6	9.5	8.8	9	-0.2
126	9	9	9	8	8.5	9.5	8.8	8.9	-0.1
127	3	1	0	0	0	0	1.1	1	0.1
128	7.5	8	8.6	7.5	9	8	8.6	8.5	0.1
129	5	6.5	7	6	7.5	7	7.8	7.6	0.2
130	10	9	9	10	10	10	10	10	0
131	5	4	6	6.6	5.9	4	4	6.3	-2.3
132	2	1	2	0	0	0	1.1	1	0.1
133	6.5	5	5	4.5	6.5	6	7.2	6.3	0.9
134	5	4.5	4	4.9	5	5.5	3.9	5.1	-1.2
135	6	7.5	7	6.5	7	6.5	7.8	7.6	0.2
136	8.5	8	8.5	8	9	9.5	8.8	9	-0.2
137	7.8	7	6.9	5	6.7	5	7.2	7.6	-0.4
138	3	2	1	1	1	0	1.1	1	0.1
139	9	9.5	10	10	9.5	10	10	10	0
140	5.9	6	7.5	7	8	7	7.8	7.6	0.2
141	8	8.5	8	7.5	8	8.5	8.6	8.9	-0.3
142	9	8.9	9	9	8.5	9	8.8	9.1	-0.3
143	8	8.5	7	7.5	8	7.5	7.9	7.7	0.2

Con esos datos que arrojaron ambos modelos matemáticos difusos se procedió a calcular estadísticos descriptivos con ayuda del *software* Minitab, los resultados se aprecian en la Tabla 6.

**Tabla 6. Estadísticas descriptivas de ambos modelos difusos**

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	CoefVar	Mínimo	Mediana	Máximo
Modelo									
Predecesor	143	0	7.457	0.188	2.253	30.22	0	7.8	10
SEEI	143	0	7.455	0.189	2.256	30.26	0	7.7	10

Al observar los estadísticos podemos percatarnos que no hubo cambios significativos (una unidad o más), en la media o promedio, se mantiene en 7.45 de calificación con ambos instrumentos, esto significa que hay evidencia, cuando menos con este muestreo, que el promedio general no se ve alterado al utilizar el *software* predecesor o SEEI.

En cuanto a la dispersión de la frecuencia de los datos (calificaciones), también encontramos que la desviación estándar prácticamente es la misma en ambos modelos, lo que sugiere que la distribución de las calificaciones respecto a la media, con ambos instrumentos están prácticamente dispersas de forma similar para los dos *softwares*.

Esto último nos sugiere que ambos instrumentos tienen aproximadamente la misma precisión, sin embargo, al poderse clasificar en varias muestras, dos en total (una por cada instrumento) la desviación estándar puede ser engañosa, ya que no nos permite conocer la precisión de un instrumento (muestra) respecto al otro. Por esta razón que se debe interpretar el coeficiente de variación  $\text{CoefVar} = \text{Desviación estándar}/\text{media}$ , que es una medida de dispersión relativa, porque mide la precisión de los datos con respecto a la media, y al ser un coeficiente, el resultado se interpreta a manera porcentual.

Al interpretar el coeficiente de variación se encuentra que el modelo difuso predecesor tuvo un porcentaje de variación del 30.22% y SEEI del 30.26%, siendo prácticamente igual de precisos para este muestreo, dejando al predecesor con la delantera con tan solo 0.04% más precisión, por lo que la primera hipótesis planteada se acepta como válida ya que hay evidencia que: Ambos modelos difusos conservan una precisión similar, con una diferencia menor a 5 unidades porcentuales.

En lo que respecta a la segunda hipótesis planteada, que hace referencia a la igualdad de promedios con los dos instrumentos, no basta con basarnos en el resultado de la Tabla 6. Fue necesario realizar una prueba de hipótesis ANOVA con Test de Tukey para poder inferir. Los resultados que arrojó Minitab para ANOVA se muestran en la Tabla 7.

**Tabla 7. Resultados del análisis de varianza ANOVA para ambos instrumentos**

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Instrumento	1	0	0.00014	0	0.996
Error	284	1443.48	5.08269		
Total	285	1443.48			

De los resultados de la Tabla 7, el Valor p es mayor a 0.05, es decir, es mayor al 5% de nivel de significancia, lo que da evidencia para aceptar la hipótesis estadística  $H_0$  como válida, esta evidencia se vuelve más fuerte al revisar los resultados del Test de Tukey mostrados en la Tabla 8.

**Tabla 8. Agrupar información utilizando el método de Tukey y una confianza de 95%**

Instrumento	N	Media	Agrupación
DIFUSO	143	7.457	A
SEEI	143	7.455	A

La última columna Agrupación revela que, al pertenecer ambos al mismo grupo A, existe evidencia que la media o promedio no cambia al evaluar a los estudiantes con el *software* predecesor o con SEEI.

Por los resultados arrojados con ANOVA y Test de Tukey se determina que la hipótesis estadística  $H_0: (\mu_1 = \mu_2)$  es válida, por lo tanto, no hay diferencia en las medias de las mediciones calculadas con los dos instrumentos.

## DISCUSIÓN Y CONCLUSIONES

Existe una alta similitud entre los resultados arrojados por SEEI y los arrojados por el modelo difuso presentado por Sánchez et al. (2020), que en realidad son los resultados del *software* predecesor. En esa investigación se mostró cómo el modelo difuso resultó ser más preciso que las rúbricas contra las que se comparó, y a su vez, la presente investigación reveló una precisión similar en SEEI, lo que permite deducir que SEEI fue también mejor que las rúbricas en términos de precisión matemática.

Así pues, los resultados de ambos *softwares* resultaron ser igual de precisos, además el promedio general de todos los alumnos evaluados no se vio alterado, estas dos cualidades dan validez a SEEI y brindan evidencia suficiente para que este *software* siga evolucionando, ya que se reafirma el poder de la Inteligencia Artificial sobre los modelos matemáticos tradicionales a base de escalas.

Se recomienda entonces realizar un nuevo estudio para probar a SEEI en otros contextos, es decir, al evaluar otras competencias de otras asignaturas para volver a poner a prueba su validez en situaciones más exigentes.

## REFERENCIAS BIBLIOGRÁFICAS

- Cano, Elena (2015). LAS RÚBRICAS COMO INSTRUMENTO DE EVALUACIÓN DE COMPETENCIAS EN EDUCACIÓN SUPERIOR: ¿USO O ABUSO?. Profesorado. Revista de Currículum y Formación de Profesorado, 19(2),265-280.[fecha de Consulta 5 de Septiembre de 2022]. ISSN: 1138-414X. Disponible en: <https://www.redalyc.org/articulo.oa?id=56741181017>.
- Del Canto, Ero, y Silva Silva, Alicia (2013). Metodología cuantitativa: abordaje desde la complementariedad en ciencias sociales. Revista de Ciencias Sociales (Cr), III (141), pp. 25-34.
- Gutiérrez, A. & Ferreira, W. (2020). Un modelo de regresión lineal aplicando lógica difusa. Revista Sextante, 23, pp. 48 - 54, 2020.
- Hernández, R., Fernández, C., y Baptista, M. (2010). Metodología de la investigación (5th ed.). México D.F.: McGRAW-HILL.
- Jamsandekar, S., y Mudholkar, R. (2013). Performance Evaluation by Fuzzy Inference Technique. International Journal Of Soft Computing And Engineering, 3(2), pp. 158-164.
- Martínez, J. (2008). Las rúbricas en la evaluación escolar: su construcción y uso. Avances en Medición, 6, 129-138.
- Monje, C. (2011). Metodología de la investigación cuantitativa y cualitativa. Guía didáctica. Neiva: Universidad sur colombiana.
- Picón Jácome, Édgar (2013). La rúbrica y la justicia en la evaluación. Íkala, revista de lenguaje y cultura, 18(3),79-94.[fecha de Consulta 5 de Septiembre de 2022]. ISSN: 0123-3432. Disponible en: <https://www.redalyc.org/articulo.oa?id=255030038006>.
- Sánchez García, J. E. ., Valdez Sandoval, A. ., Soto Vega, J. E. ., & Gutiérrez Herrera, B. E. . (2021). Comparación del nivel de desempeño de una competencia usando tres instrumentos, dos basados en rúbrica y otro basado en lógica difusa: A comparison of the level of competency using three instruments; two rubric based instruments and a fussy logic-based instrument. Revista Relep - Educación Y Pedagogía En Latinoamérica, 2(4), 123–145. <https://doi.org/10.46990/relep.2020.2.4.245>.