

1. EVALUACIÓN DE LOS APRENDIZAJES CON INTELIGENCIA ARTIFICIAL EN EDUCACIÓN MEDIA SUPERIOR

Evaluation of Learning with Artificial Intelligence in Higher Secondary Education

Alberto Valdez Sandoval¹, Gladis Andrea Soto López²

Fecha recibido: 20/11/2022

Fecha aprobado: 15/12/2022

Derivado del proyecto: Sistema Experto de Evaluación Inteligente.

Institución financiadora: Financiación propia.

Pares evaluadores: Red de Investigación en Educación, Empresa y Sociedad – REDIEES³.

¹ 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.

² Ingeniero Bioquímico, Instituto Tecnológico de Culiacán, Ocupación (Docente), Centro de Bachillerato Tecnológico industrial y de servicios CBTis 224, correo electrónico: gladis.soto@cbtis224.edu.mx.

³ Red de Investigación en Educación, Empresa y Sociedad – REDIEES.

RESUMEN

En el presente trabajo se comparan los resultados al medir aprendizajes de estudiantes de educación media superior en las asignaturas de física y química, en CBTis 224. La comparación se lleva a cabo midiendo los aprendizajes arrojados por dos instrumentos: a) el Sistema Experto de Evaluación Inteligente (SEEI) publicado por Sánchez et al. (2021) y b) Una evaluación tradicional utilizando ponderaciones. SEEI es un sistema experto que utiliza 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. Al utilizar IA se espera que los resultados arrojados sean más precisos y exactos que las evaluaciones tradicionales con por ponderaciones ya que la IA imita el razonamiento humano que sucede en el cerebro del docente cuando evalúa y emite un juicio de valor. El objetivo de esta investigación es encontrar validez en el instrumento SEEI, se busca con esta línea de investigación encontrar de manera paulatina esta validez que le permita 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.

PALABRAS CLAVE: *lógica difusa; sistema experto; evaluación; competencia; rúbrica.*

ABSTRACT

In the present research work, the results are compared when measuring the learning of high school students in the subjects of physics and chemistry, in CBTis 224. The comparison is carried out by measuring the learning produced by two instruments: a) the Expert Evaluation System Intelligent (SEEI) published by Sánchez et al. (2021) and b) A traditional evaluation using estimation. SEEI is an expert system that uses Artificial Intelligence (AI) to calculate, on a scale from zero to ten, the grade resulting from a student learning assessment process. By using AI, it is expected that the results obtained will be more precise and exact than traditional evaluations with weights, since AI imitates the human reasoning that occurs in the teacher's brain when evaluating and making a value judgment. The objective of this research is to find validity in the SEEI instrument, it is sought with this line of research to gradually find this validity that allows it to be used by any teacher, to evaluate any competence, in any subject, in any educational institution located in anyplace in the world just by having internet access. The results provide firm evidence that the line of research is on the right track and that it is worth continuing testing and adjusting SEEI and finding validity in more contexts in the near future.

KEYWORDS: *fuzzy logic; expert system; assessment; competency; rubric.*

INTRODUCCIÓN

La Lógica Difusa (LD) es una rama de la Inteligencia Artificial (IA) que permite crear modelos matemáticos y procesarlos mediante un computador. Estos modelos matemáticos consisten en imitar el razonamiento humano cuando se toman decisiones en las que existe incertidumbre por la subjetividad.

Un ejemplo de esta toma de decisiones es cuando un docente evalúa el desempeño de un estudiante, ya que en este proceso mental tan complejo y lleno de subjetividad el docente debe generar un juicio de valor que sea justo para medir el progreso del estudiante.

Gutiérrez y Ferreira (2020) visualizan 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 permite esta imprecisión, se obtienen resultados indeseables (p.2).

Actualmente la mayoría de profesionistas, a excepción de los docentes, cuentan con instrumentos de medición digitales de alta precisión y exactitud que les permite realizar sus labores de forma más adecuada, como ejemplos podemos mencionar: a) un piloto aviador, tiene instalados en su aeronave instrumentos para medir la velocidad y altura de manera eficiente, b) un ingeniero de *software* cuenta con instrumentos que le permitan medir la eficiencia del *software* que está desarrollando, c) los entrenadores de *pitchers* de *baseball* cuentan con pistolas de efecto Doppler para medir la velocidad de lanzamiento de la bola.

Los docentes por su parte cuentan con instrumentos de medición de aprendizajes obsoletos y con modelos matemáticos que no son adecuados para medir el aprendizaje. Estos modelos matemáticos consisten en promedios o ponderaciones (regla de 3) que distan bastante del proceso que sucede dentro del cerebro del docente cuando este está evaluando y emitiendo un juicio de valor.

Esta problemática provoca que los docentes, una vez que tienen las calificaciones calculadas decidan hacer ajustes, por ejemplo, un alumno que alcanzó un 9.3 pudiera quizás

ser considerado un alumno de 10, esto sucede porque el modelo matemático utilizado no refleja el juicio de valor que emitió su cerebro.

En el estudio previo 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.

En ese estudio el *software* se desarrolló exclusivamente para probabilidad y estadística, sin embargo, la actualización del mismo se llamó Sistema Experto de Evaluación Inteligente (SEEI), el cual fue adaptado para poder ser utilizado en cualquier asignatura, de cualquier nivel educativo, en cualquier institución que se encuentre en cualquier parte del mundo, es decir, el modelo difuso es adaptativo (sin que el usuario lo note).

En la presente investigación se busca seguir dando validez a SEEI en otro contexto, para ello, a diferencia del estudio de Sánchez et al. (2021), aquí las condiciones cambiaron en los sentidos: a) este estudio se llevó a cabo en otro plantel educativo, y con la participación de una docente distinta, en CBTis 224, b) se llevó a cabo en otra ciudad, c) se evaluaron otras asignaturas, física ii y química i, por lo tanto se evalúan otras competencias, d) con grupos de especialidades distintas ya que es un bachillerato tecnológico.

El propósito de esta investigación es comparar los resultados de los modelos matemáticos utilizados por la docente en las asignaturas de física ii y química i contra los resultados que arroja SEEI, para determinar cuál de los modelos fue más preciso y exacto.

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, este se llevó a cabo con cuatro grupos, con un tamaño de muestra $n=187$ estudiantes, todos evaluados con un modelo de regla de 3 por la misma docente, como se observa en la Tabla 1.

Tabla 1

Grupos muestreados

Grupo	Número de alumnos	Especialidad	Tuno	Asignatura	Competencia a evaluar
1F	51	Programación	Matutino	Química I	Obtiene, registra y sistematiza la información para responder a preguntas de carácter científico, consultando fuentes relevantes y realizando experimentos pertinentes.
5A	51	Producción industrial de alimentos	Matutino	Física II	Diseña modelos o prototipos para resolver problemas, satisfacer necesidades o demostrar principios científicos.
5B	51	Producción industrial de alimentos	Matutino	Física II	Diseña modelos o prototipos para resolver problemas, satisfacer necesidades o demostrar principios científicos.
5L	34	Mecánica industrial	Vespertino	Física II	Diseña modelos o prototipos para resolver problemas, satisfacer necesidades o demostrar principios científicos.

Nota: elaboración propia

En la Tabla 2 se puede apreciar cuales fueron los modelos matemáticos utilizados para cada asignatura/competencia.

Tabla 2

Modelos matemáticos tradicionales que se utilizaron por la docente.

Asignatura	Criterio	Dimensión de competencia	Ponderación	Modelo matemático
Química I	Tareas/ejercicios	Teoría	25%	Calificación = (Tareas*0.25) + (Examen*0.35) + (Prácticas*0.10) + (Disciplina*0.10) + (Proyecto*0.20)
	Examen	Teoría	35%	
	Práctica	Práctica	10%	
	Disciplina	Actitud	10%	
	Proyecto	Práctica	20%	
Física II	Tareas/ejercicios	Teoría	40%	Calificación = (Tareas*0.40) + (Prácticas*0.10) + (Disciplina*0.10) + (Proyecto*0.40)
	Práctica	Práctica	10%	
	Disciplina	Actitud	10%	
	Proyecto	Práctica	40%	

Nota: elaboración propia.

El procedimiento para realizar la comparación consistió en que la docente evaluó a sus estudiantes con los modelos matemáticos antes mencionados en la Tabla 2 y una vez que tenía calificaciones por cada criterio, esas mismas calificaciones por criterios se ingresaron a SEEI para que también calculara una calificación resultante y así poder comparar la calificación emitida del modelo matemático tradicional contra la emitida por SEEI, esto para cada estudiante.

La docente verificó ambas calificaciones (su modelo y la que arrojó SEEI) para cada estudiante y verificó por cada uno de ellos cuál de los dos resultados se apegaba más a su juicio de valor, es decir, cual resultado fue más exacto.

En lo que respecta al modelo matemático de SEEI, este es adaptativo, la Tabla 3 expone las funciones de membresía que se utilizan.

Tabla 3

Funciones de membresía de entrada del modelo difuso de SEEI

ENTRADAS/ATRIBUTOS	FUNCIONES DE MEMBRESÍA
Teoría	Bajo: $y = \text{trimf}(x, [0 \ 0 \ 4])$ Medio: $y = \text{trapmf}(x, [3 \ 5 \ 6.5 \ 7])$ Alto: $y = \text{gaussmf}(x, [1.5 \ 10])$
Práctica	Bajo: $y = \text{trimf}(x, [0 \ 0 \ 3])$ Medio: $y = \text{trapmf}(x, [2 \ 5 \ 6 \ 8])$ Alto: $y = \text{gaussmf}(x, [1.699 \ 10])$
Actitud	Bajo: $\text{trimf}(x, [0 \ 0 \ 3])$ Regular: $\text{trimf}(x, [1.1672 \ 3.672 \ 5.672])$ Bueno: $\text{gbellmf}(x, [1.05 \ 1.66 \ 6.678])$ Alto: $y = \text{gaussmf}(x, [1 \ 9.964])$

Nota: elaboración propia

Una vez que se obtuvieron los resultados entregados por SEEI, se dio paso a someter a prueba las siguientes hipótesis para el estudio 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:

H₀: ($\mu_1 = \mu_2$). No hay diferencia en las medias de las mediciones calculadas con los dos instrumentos.

H₁: ($\mu_1 \neq \mu_2$). 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 generados por los modelos matemáticos tradicionales y SEEI se aprecian en las Tablas 4 y Tabla 5, específicamente en las últimas dos columnas. Las calificaciones marcadas en color gris son las que la docente identificó que el modelo tradicional fue más exacto.

Tabla 4

Resultados de la asignatura de Química I para el grupo 1F de programación

ALUMNO	TEORÍA-25% Tareas/ejercicios	TEORÍA-35% Examen	PRACTICA-10% Práctica	ACTITUD-10% Disciplina	PRACTICA-20% Proyecto	CALIF TRADICIONAL	CALIF SEEI
1	8.6	7.8	8.5	9.0	0.0	6.6	7.0
2	0.0	2.6	0.0	8.0	10.0	3.7	5.0
3	9.5	7.0	10.0	10.0	10.0	8.8	9.5
4	9.6	7.8	8.5	10.0	6.2	8.2	8.5
5	8.0	4.8	5.0	10.0	10.0	7.2	7.7
6	7.0	4.1	9.0	9.0	10.0	0.0	7.7
7	5.3	8.5	5.0	10.0	6.2	7.0	7.8
8	10.0	7.0	8.5	9.0	10.0	8.7	8.5
9	3.2	6.3	5.0	8.0	10.0	6.3	7.5
10	9.6	9.6	9.5	10.0	10.0	9.7	10.0
11	6.0	4.8	5.0	5.0	10.0	6.2	7.7
12	9.0	3.0	9.5	10.0	10.0	7.2	8.5
13	8.8	5.9	9.0	5.0	10.0	7.7	7.7
14	4.9	6.3	5.0	5.0	10.0	6.4	7.7
15	8.3	3.7	8.5	10.0	6.2	6.5	7.6
16	6.8	5.9	5.0	9.0	10.0	7.2	7.8
17	9.6	9.6	8.5	10.0	6.2	8.9	8.5
18	8.6	5.9	5.0	8.0	10.0	7.5	7.7
19	0.0	3.7	0.0	10.0	0.0	2.3	3.5
20	7.1	6.3	8.0	9.0	10.0	7.7	8.6
21	0.0	1.5	5.0	8.0	10.0	3.8	5.0
22	10.0	9.3	10.0	10.0	10.0	9.7	10.0
23	9.9	8.9	10.0	10.0	10.0	9.6	10.0
24	8.8	5.6	5.0	10.0	10.0	7.6	7.8
25	9.4	7.4	5.0	8.0	10.0	8.2	7.5
26	9.0	10.0	9.0	10.0	10.0	9.7	10.0
27	10.0	8.1	10.0	10.0	10.0	9.4	10.0

28	9.3	5.6	5.0	8.0	10.0	7.6	7.7
29	3.5	4.1	5.0	10.0	10.0	5.8	7.6
30	7.8	7.8	9.5	10.0	10.0	8.6	10.0
31	0.0	0.4	0.0	7.0	0.0	0.8	2.0
32	0.0	3.0	5.0	10.0	6.2	3.8	5.0
33	7.2	8.9	2.5	9.0	10.0	8.1	7.5
34	4.2	5.6	3.5	10.0	0.0	4.3	5.0
35	9.8	9.6	8.5	10.0	10.0	9.7	10.0
36	9.4	10.0	9.0	10.0	10.0	9.8	10.0
37	7.5	8.1	5.0	10.0	6.2	7.5	7.6
38	0.0	4.8	0.0	10.0	6.2	3.9	5.0
39	2.7	4.4	5.0	9.0	10.0	5.6	7.6
40	8.0	8.9	9.0	10.0	10.0	9.0	10.0
41	4.3	7.4	5.0	9.0	10.0	7.1	7.5
42	8.2	4.4	8.5	10.0	10.0	7.5	8.6
43	7.2	5.2	5.0	7.0	10.0	6.8	7.6
44	3.0	4.8	5.0	9.0	6.2	5.1	6.3
45	8.6	3.7	5.0	8.0	6.2	6.0	7.6
46	9.6	9.6	5.0	10.0	10.0	9.3	8.5
47	9.3	8.1	8.5	10.0	10.0	9.0	10.0
48	8.8	9.3	9.5	10.0	6.2	8.6	8.5
49	8.1	2.2	9.5	10.0	10.0	6.8	7.7
50	6.2	3.7	8.5	10.0	10.0	6.7	7.6
51	9.0	4.8	8.5	9.0	6.2	6.9	7.8

Nota: elaboración propia

Tabla 5

Resultados de la asignatura de Física II para los grupos 5A y 5B de alimentos y 5L de mecánica

ALUMNO	TEORÍA-40%	PRÁCTICA-10%	ACTITUD-10%	PRÁCTICA-40%	CALIF TRADICIONAL	CALIF SEEI
	Tareas/ejercicios	Práctica	Disciplina	Proyecto		
52	9.6	9.3	9.0	9.5	9.5	10.0
53	9.3	9.5	9.0	8.8	9.1	10.0
54	9.4	0.0	10.0	8.8	8.3	7.0
55	6.9	9.5	10.0	9.8	8.6	9.4
56	9.8	9.3	9.0	9.3	9.4	10.0
57	9.9	9.8	9.0	8.5	9.2	10.0
58	9.9	9.3	10.0	9.3	9.6	10.0
59	9.9	9.5	10.0	9.5	9.7	10.0

60	9.8	9.3	10.0	9.3	9.5	10.0
61	9.9	9.8	8.0	9.3	9.4	9.5
62	9.1	9.3	9.0	8.8	9.0	10.0
63	9.0	9.3	8.0	9.5	9.1	8.5
64	9.5	0.0	10.0	9.3	8.5	7.0
65	9.6	9.3	8.0	9.5	9.4	9.5
66	5.0	9.3	9.0	9.3	7.5	8.5
67	6.8	5.0	7.0	9.5	7.7	7.8
68	9.9	9.8	9.0	8.5	9.2	10.0
69	9.9	9.3	10.0	9.3	9.6	10.0
70	9.5	9.3	10.0	9.5	9.5	10.0
71	10.0	9.3	9.0	9.3	9.5	10.0
72	9.6	9.3	10.0	9.3	9.5	10.0
73	9.6	9.8	8.0	9.3	9.3	9.5
74	9.5	9.5	8.0	9.0	9.2	9.5
75	9.8	9.5	9.0	9.5	9.6	10.0
76	9.8	9.3	9.0	8.8	9.2	10.0
77	9.9	9.3	10.0	9.5	9.7	10.0
78	9.8	9.3	10.0	9.3	9.5	10.0
79	9.8	9.3	9.0	9.3	9.4	10.0
80	9.8	9.5	9.0	9.3	9.5	10.0
81	9.5	9.5	9.0	9.0	9.3	10.0
82	2.5	9.3	9.0	9.3	6.5	7.7
83	9.5	9.5	9.0	9.0	9.3	10.0
84	9.6	9.3	10.0	8.5	9.2	10.0
85	9.5	9.8	10.0	9.3	9.5	10.0
86	8.9	9.8	8.0	9.5	9.1	9.5
87	9.5	9.5	9.0	9.3	9.4	10.0
88	9.6	9.5	8.0	9.0	9.2	9.5
89	9.8	9.5	9.0	9.0	9.4	10.0
90	9.8	9.3	9.0	9.3	9.4	10.0
91	9.9	9.8	8.0	9.0	9.3	9.5
92	8.0	9.3	10.0	8.3	8.4	9.0
93	10.0	9.3	10.0	8.5	9.3	10.0
94	10.0	9.8	9.0	9.5	9.7	10.0
95	9.8	9.8	8.0	9.3	9.4	9.5
96	9.5	9.5	8.0	9.3	9.3	9.5
97	9.6	9.5	8.0	9.3	9.3	9.5
98	7.4	9.3	9.0	8.3	8.1	8.8
99	9.8	9.8	8.0	9.3	9.4	9.5
100	9.9	9.3	9.0	9.5	9.6	10.0
101	7.4	9.8	9.0	9.0	8.4	8.8
102	9.9	9.3	10.0	9.5	9.7	10.0

103	9.6	10.0	9.0	9.5	9.6	10.0
104	4.6	10.0	9.0	9.0	7.4	8.5
105	9.6	9.5	9.0	9.0	9.3	10.0
106	8.9	9.5	9.0	9.0	9.0	10.0
107	9.6	9.8	10.0	9.8	9.7	10.0
108	7.5	9.5	7.0	8.3	8.0	8.5
109	9.5	10.0	7.0	9.0	9.1	8.5
110	8.8	9.5	8.0	8.8	8.8	8.5
111	2.5	9.5	9.0	9.3	6.6	7.7
112	9.6	9.5	9.0	9.0	9.3	10.0
113	0.0	0.0	0.0	0.0	0.0	0.0
114	7.0	10.0	7.0	9.0	8.1	8.5
115	6.8	9.5	9.0	9.0	8.2	8.5
116	9.3	9.5	8.0	8.5	8.9	8.5
117	9.9	9.8	10.0	9.3	9.6	10.0
118	9.3	9.5	9.0	9.3	9.3	10.0
119	9.9	10.0	9.0	9.8	9.8	10.0
120	9.6	9.8	9.0	9.5	9.5	10.0
121	9.3	4.5	10.0	9.0	8.8	8.5
122	9.9	9.8	7.0	8.0	8.8	8.5
123	9.6	10.0	9.0	9.8	9.7	10.0
124	9.5	10.0	7.0	9.0	9.1	8.5
125	9.6	10.0	8.0	9.3	9.4	9.5
126	9.3	9.5	10.0	9.0	9.3	10.0
127	8.1	9.5	9.0	9.0	8.7	9.0
128	7.1	10.0	7.0	9.3	8.3	8.6
129	9.5	9.5	7.0	8.5	8.9	8.5
130	9.4	9.5	8.0	9.0	9.1	8.5
131	7.5	9.5	7.0	9.5	8.5	8.5
132	9.6	9.5	7.0	9.0	9.1	8.5
133	9.8	10.0	8.0	8.3	9.0	9.5
134	9.8	9.5	9.0	9.5	9.6	10.0
135	9.5	9.5	7.0	9.3	9.2	8.5
136	9.9	10.0	10.0	9.5	9.8	10.0
137	9.8	9.8	10.0	9.5	9.7	10.0
138	10.0	10.0	10.0	8.5	9.4	10.0
139	9.9	9.5	10.0	8.8	9.4	10.0
140	9.6	9.5	8.0	8.3	8.9	8.5
141	6.3	9.5	8.0	8.5	7.7	8.5
142	9.6	10.0	10.0	9.0	9.5	10.0
143	9.4	9.5	10.0	8.5	9.1	10.0
144	9.3	10.0	9.0	9.3	9.3	10.0
145	5.0	9.5	8.0	8.3	7.1	7.7

146	9.6	8.5	10.0	9.3	9.4	10.0
147	8.8	9.5	9.0	8.5	8.8	9.1
148	5.0	9.5	10.0	8.5	7.4	8.5
149	9.6	9.5	10.0	9.3	9.5	10.0
150	4.3	10.0	8.0	8.5	6.9	7.7
151	9.9	9.8	8.0	8.3	9.0	9.5
152	9.5	10.0	9.0	9.5	9.5	10.0
153	4.8	9.8	8.0	9.3	7.4	7.7
154	0.0	0.0	0.0	0.0	0.0	0.0
155	7.5	10.0	10.0	8.5	8.4	10.0
156	5.3	8.5	9.0	8.5	7.3	7.8
157	0.0	0.0	9.0	8.0	4.1	5.3
158	0.0	0.0	8.0	8.0	4.0	5.0
159	6.5	9.0	10.0	8.0	7.7	8.5
160	6.5	10.0	9.0	7.5	7.5	8.5
161	9.3	7.5	8.0	7.8	8.4	8.5
162	4.8	8.5	8.0	7.5	6.6	7.6
163	2.5	10.0	10.0	8.5	6.4	7.7
164	4.8	8.5	9.0	8.3	7.0	7.8
165	2.3	0.0	8.0	9.0	5.3	5.0
166	7.0	0.0	10.0	8.3	7.1	7.0
167	0.0	0.0	8.0	8.0	4.0	5.0
168	8.8	9.5	10.0	8.5	8.9	10.0
169	8.0	7.5	8.0	8.5	8.2	8.5
170	0.0	7.5	8.0	7.0	4.4	5.0
171	2.5	9.0	9.0	8.8	6.3	7.7
172	2.5	8.5	8.0	8.0	5.9	7.7
173	2.5	7.5	9.0	8.5	6.1	7.6
174	1.8	7.5	8.0	8.5	5.7	7.6
175	0.0	8.5	9.0	8.8	5.3	7.0
176	6.8	8.5	10.0	8.0	7.8	8.5
177	5.0	7.5	8.0	7.8	6.7	7.6
178	5.0	0.0	9.0	8.5	6.3	5.0
179	9.5	7.5	9.0	8.5	8.9	8.7
180	9.8	9.0	8.0	8.8	9.1	8.5
181	9.0	10.0	9.0	8.8	9.0	10.0
182	5.0	0.0	7.0	8.8	6.2	5.0
183	2.5	7.5	8.0	8.5	6.0	7.6
184	0.0	0.0	8.0	8.0	4.0	5.0
185	5.0	9.0	9.0	7.5	6.8	7.6
186	7.5	9.0	8.0	8.8	8.2	8.5
187	5.0	9.0	8.0	8.5	7.1	7.7

Nota: elaboración propia.

Una vez obtenidos los resultados de las calificaciones que arrojaron los instrumentos, se procedió a analizarlos estadísticamente con ayuda del *software* Minitab. La Tabla 6 evidencia las estadísticas descriptivas.

Tabla 6

Comparación de estadísticas descriptivas de ambos resultados

Variable	Media	Desv.Est.	Mínimo	Máximo	Modo	N para moda
TRADICIONAL	7.938	1.966	0	9.75	9.25, 9.525	5
SEEI	8.492	1.761	0	10	10	65

Nota: elaboración propia.

Al analizar los resultados estadísticos arrojados por Minitab podemos darnos cuenta de varios aspectos relevantes tales como: a) la media parece haber cambiado significativamente una al utilizar ambos instrumentos, b) la desviación estándar también es similar, lo que sugiere que la precisión en las calificaciones arrojadas es similar con ambos métodos, c) con los instrumentos tradicionales ningún estudiante alcanzó la calificación de 10, d) el valor que más se repite en las calificaciones arrojadas por SEEI es 10.0, mientras como ya se mencionó, con el procedimiento tradicional ningún estudiante alcanzó esta calificación.

También podemos observar en la Tabla 4 y Tabla 5, los valores marcados en color gris. Estos representan las calificaciones que la docente considera que el modelo tradicional fue más exacto (más justo), en total son 38 calificaciones de 187 estudiantes, esto significa que el 20.3% de los resultados el sistema tradicional fue más exacto/justo y 149 calificaciones que equivale al 79.7% fue más exacto/justo SEEI.

Por otra parte, regresando a la Tabla 6, aunque la desviación estándar sugiere que ambos instrumentos muestran una distribución (precisión) similar, se requiere del cálculo de coeficiente de variación para determinar si la hipótesis de diferencia entre grupos es aceptada como válida:

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

El coeficiente de variación 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. Minitab arroja los siguientes resultados, mostrados en la Tabla 7.

Tabla 7

Coeficiente de variación de ambos métodos de evaluación

Variable	CoefVar
TRADICIONAL	24.76
SEEI	20.74

Nota: elaboración propia.

Al interpretar el coeficiente de variación se encuentra que SEEI tuvo un porcentaje de variación del 20.74% y el método tradicional 24.76%, siendo más preciso SEEI por 4.02% para este muestreo.

Por lo que existe evidencia para afirmar que la hipótesis H_i. Ambos modelos difusos conservan una precisión similar, con una diferencia menor a 5 unidades porcentuales, es tomada como verdadera.

Ya evaluada la precisión de los instrumentos se realizó un análisis de varianza ANOVA y Test de Tukey para someter a prueba la hipótesis estadística:

H₀: ($\mu_1 = \mu_2$). No hay diferencia en las medias de las mediciones calculadas con los dos instrumentos.

H₁: ($\mu_1 \neq \mu_2$). Existe diferencia en las medias de las mediciones calculadas con los dos instrumentos.

Los resultados de Minitab se observan en la Tabla 8.

Tabla 8

Análisis de varianza ANOVA con una confianza del 95%

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
INSTRUMENTO	1	28.69	28.69	8.24	0.004
Error	372	1295.68	3.483		
Total	373	1324.37			

Nota: elaboración propia.

Dado que el Valor p es mayor a 0.05, es decir, es menor al 5% de nivel de significancia, lo que da evidencia para aceptar la hipótesis estadística H_0 como no válida, aceptando como válida H_1 (las medias o promedios grupales han cambiado con el uso de un método de evaluación y otro), esta evidencia se vuelve más fuerte al revisar los resultados del Test de Tukey mostrados en la Tabla 9.

Tabla 9

Test de Turkey con una confianza del 95%

INSTRUMENTO	N	Media	Agrupación
SEEI	187	8.492	A
TRADICIONAL	187	7.938	B

Nota: elaboración propia.

Al analizar el resultado del Test arrojado por Minitab, vemos que en la última columna Agrupación ambos instrumentos no pertenecen a la misma agrupación, sino que SEEI pertenece al grupo A y el método tradicional al grupo B, esto se interpreta como que ambas medias o promedios han cambiado significativamente al utilizar un método de evaluación u otro, reforzando la decisión de tomar como válida H_1 : las medias o promedios grupales han cambiado con el uso de un método de evaluación y otro.

DISCUSIÓN Y CONCLUSIONES

En el modelo difuso presentado por Sánchez et al. (2020), que a la vez fue el software predecesor de SEEI se observa que la media no cambió significativamente al usar un método tradicional de rúbricas, a diferencia de este estudio en donde la media si cambió con un instrumento u otro, como pudo dar evidencia el ANOVA y Test de Tukey.

En la presente investigación también existe sólida evidencia de que SEEI es más preciso y exacto que los métodos tradicionales aquí empleados, ya que SEEI fue, según el coeficiente de variación 4.02% más preciso para este muestreo.

En lo que respecta a la exactitud, SEEI también salió vencedor, pues la docente consideró que fue 79.7% fue más exacto/justo que la evaluación tradicional.

Finalmente pudimos observar que con SEEI algunos de los estudiantes si alcanzaron la calificación máxima de 10.0, esto es porque los docentes en muchos de los casos, aunque los instrumentos no arrojen una calificación de 10, aun así consideramos que los estudiantes ya han alcanzado el nivel de logro de la competencia planteado, simplemente el instrumento obsoleto que usamos no alcanza a captar la complejidad del modelo matemático que sucede dentro del cerebro humano.

Con esto se reafirma el poder de la Inteligencia Artificial sobre los métodos tradicionales (obsoletos) que se utilizan diariamente en el aula y que su uso orilla a los docentes a dar ajustes a las calificaciones finales.

Se recomienda, por lo cual, realizar un nuevo estudio para poner a prueba al SEEI en otros entornos, es decir, al evaluar otras competencias de diferentes asignaturas para volver a testear su validez en situaciones más rigurosas.

REFERENCIAS BIBLIOGRÁFICAS

- Del Canto, E. y Silva Silva, A. (2013). Metodología cuantitativa: abordaje desde la complementariedad en ciencias sociales. *Revista de Ciencias Sociales*, 3(141), 25-34.
- Gutiérrez, A. y Ferreira, W. (2020). Un modelo de regresión lineal aplicando lógica difusa. *Revista Sextante*, 23, pp. 48-54. <https://doi.org/10.54606/Sextante2020.v23.05>
- Hernández, R., Fernández, C., y Baptista, M. (2010). Metodología de la investigación (5^a ed.). McGraw-HILL.
- Monje, C. (2011). *Metodología de la investigación cuantitativa y cualitativa. Guía didáctica*. Neiva: Universidad Sur Colombiana.
- Sánchez García, J. E., Valdez Sandoval, A., Soto Vega, J. E. y 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. *Revista Relep - Educación Y Pedagogía En Latinoamérica*, 2(4), 123-145. <https://doi.org/10.46990/relep.2020.2.4.245>.