



UNIVERSIDAD DE JAÉN

**FACULTAD DE HUMANIDADES Y CIENCIA
DE LA EDUCACIÓN
DEPARTAMENTO DE PEDAGOGÍA**

TESIS DOCTORAL

**FACTORES QUE AFECTAN A LA VALIDEZ DE
LA AUTOEVALUACIÓN Y SUS IMPLICACIONES
EN LA PRÁCTICA DOCENTE**

PRESENTADA POR:

DR. SAMUEL PARRA LEÓN

DIRIGIDA POR:

DR. ANTONIO PANTOJA VALLEJO

JAÉN, 04 DE Octubre DE 2021

ISBN



Universidad de Jaén

Área de Métodos de Investigación en Educación
Departamento de Pedagogía
Facultad de Humanidades y Ciencia de la Educación
UNIVERSIDAD DE JAÉN

Factores que afectan a la validez de la autoevaluación y sus implicaciones en la práctica docente

Dr. Samuel P. León

Tesis Doctoral

Doctorado en Innovación Didáctica y Formación de Profesorado

Dirigida por

Dr. Antonio Pantoja Vallejo

Índice

Agradecimiento	4
Resumen	6
Capítulo 1: EVALUACIÓN EDUCATIVA	8
1.1.- Introducción a la Evaluación Educativa	8
1.2.- Conceptualización de Evaluación Educativa	9
1.3.- Características de la Evaluación Educativa	13
Capítulo 2 AUTOEVALUACIÓN	15
2.1.-Introducción a la Autoevaluación	15
2.2.- Conceptualización de la Autoevaluación	17
2.3.-Investigación actual sobre la Autoevaluación	23
2.3.1.-Percepción de los alumnos	23
2.3.2.-Aprendizaje Auto-Regulado.....	26
2.3.3.-Validez/Precisión/Consistencia.....	29
2.4.-Implicaciones de la AE para la práctica educativa	35
Capítulo 3: INVESTIGACIÓN EXPERIMENTAL	37
3.1.- Objetivos	38
3.2.- Metodología	39
3.2.1.- Muestra	39
3.2.2.- Instrumentos	39
3.2.3.- Procedimiento	41
3.2.3.- Análisis de los datos	42
3.2.- Resultados	45
3.2.- Conclusiones	48
Bibliografía	54
Capítulo 4: ARTICULOS CIENTÍFICOS	71

Agradecimiento

Me gustaría comenzar agradeciendo al Departamento de Pedagogía de la Universidad de Jaén por darme la oportunidad y el contexto para desarrollar esta Tesis Doctoral. Especialmente me gustaría dar las gracias a mi director de tesis Antonio Pantoja Vallejo, quien en todo momento se ha mostrado accesible y dispuesto a lo largo de todo el proceso.

También me gustaría dar las gracias a todos los compañeros coautores que me han ayudado en todo este proceso. A Byron Nelson por enseñarme a implantar ese rigor científico a los trabajos, a Marta Ferrero por abrirme los ojos a la Educación Basada en Evidencias, a José María Augusto (Txefo) por toda la generosidad que desprende como persona y como investigador, y a Inma García-Martínez por ayudarme a adentrarme y comprender la investigación educativa, gracias Inma por ayudar a este rarito a sobrevivir a un mundo nuevo.

Pero si realmente tengo a alguien a los que agradecer son a mi familia. Gracias Lola y Emilio por la confianza y el apoyo a lo largo de todo este tiempo. Gracias a mis hijos Julia y Samuel, sois la energía que hace que siga en movimiento, sin duda la mejor parte de mí. Y rotundamente, gracias a ti Inma, tu apoyo, tu generosidad, tu fuerza han hecho de nuevo que esto salga adelante.

Por último, me gustaría dar las gracias a todos los familiares, amigos y compañeros que han participado directa o indirectamente en este trabajo.

A todos ellos ¡gracias!

*A Julia, Samuel e Inma,
los motores que hacen moverlo todo.*

S.P.L.

Resumen

El presente trabajo de Tesis Doctoral trata de analizar el impacto que diferentes factores propuestos tienen sobre la validez de la autoevaluación, así como sus implicaciones en la práctica educativa. Para llevar a cabo este objetivo, se realizaron cuatro estudios de investigaciones con estudiantes universitario en los que se analizó la relación o influencia de diferentes factores sobre la habilidad de los alumnos a la hora de evaluar sus logros académicos. El primero de los estudios (León, Pantoja-Vallejo y Nelson, 2021), plantea analizar la relación que existe entre la competencia académica de los alumnos y su precisión a la hora de autoevaluar sus logros académicos. En el segundo (León, Pantoja-Vallejo y Ferrero, 2021) se analiza la relación que la posición que los alumnos escogen en clase con su rendimiento académico y con la habilidad de autoevaluación de los estudiantes. El tercero de los trabajos que conforma la Tesis Doctoral (León y García-Martínez, 2021) se evalúa el efecto de la experiencia de los alumnos con la evaluación y el feedback recibido sobre la habilidad de evaluar el logro ajeno y propio. Finalmente, el cuarto trabajo (León, Augusto-Landa y García-Martínez, 2021), analiza mediante un modelo basado en ecuaciones estructurales, la relación que existe entre factores como el compromiso académico y las estrategias de estudios de los alumnos con su logro académico y su habilidad para autoevaluar dichos logros. Los resultados obtenidos nos permiten tener más conocimiento sobre la influencia que los factores propuestos tienen sobre la precisión de los alumnos a la hora de evaluar sus logros, así como analizar las implicaciones que pudieran tener sobre la práctica educativa.

Abstract

This Doctoral Thesis work tries to analyse the impact that different proposed factors have on the validity of self-assessment as well as its implications in educational practice. To achieve this objective, four research studies were carried out with university students in which the relationship or influence of different factors on the students' ability to evaluate their academic achievements was analysed. The first of the studies (León, Pantoja-Vallejo and Nelson, 2021), proposes to analyse the relationship between students' academic competence and their accuracy in self-evaluating their academic achievements. The second study (León, Pantoja-Vallejo and Ferrero, 2021) analyses the relationship that the position students choose in class has with their academic achievement and with students'

self-evaluation ability. The third of the papers included (León and García-Martínez, 2021) evaluates the effect of the students' experience with evaluation and the feedback received on the ability to evaluate the achievement of others and their own. Finally, the fourth paper analyses (León, Augusto-Landa y García-Martínez, 2021), using a model based on structural equations, the relationship between factors such as students' academic commitment and study strategies with their academic achievement and their ability to self-evaluate these achievements. The results obtained allow us to have more knowledge about the influence that the proposed factors have on the accuracy of students when evaluating their achievements, as well as to analyse the implications they could have on educational practice.

Capítulo 1: EVALUACIÓN EDUCATIVA

1.1.- Introducción a la Evaluación Educativa

En este apartado de la tesis se realizará una introducción a la evaluación educativa como proceso central en el que se basa toda evaluación, además haremos un recorrido sobre la evolución que el término evaluación ha mostrado a lo largo de la historia para dar respuesta a las necesidades educativas, revisaremos las principales funciones atribuidas a la evaluación y finalizaremos mostrando algunas de las principales clasificaciones o modalidades de evaluación educativa propuestas en la literatura. Así, este primer bloque de contenidos nos ayudará a comprender los conceptos claves necesarios para adentrarnos dentro de la modalidad de evaluación que centrará todo el trabajo de investigación de esta tesis doctoral, la autoevaluación.

La evaluación educativa es una parte esencial dentro del proceso educativo desde la constitución del mismo (Ainscow, 1988; Ysseldyke y Matson, 1988). La evaluación educativa a lo largo de la historia ha ido adquiriendo diferentes enfoques y definiciones, desde su concepción hacia el logro, propuesta por Tyler (1950), hasta enfoques más sistémicos contemporáneos (Arredondo, 2002). En la actualidad, la evaluación educativa pretende adoptar su carácter más pedagógico, entendiéndose como un proceso de asesoramiento, regulación, orientación y consolidación de lo aprendido, enfocado a la mejora de los procesos de enseñanza y aprendizaje en los contextos educativos. La importancia conferida a la evaluación en el contexto educativa y sus importantes implicaciones sobre el aprendizaje es tal que en la actualidad la evaluación es considerada un factor clave en la calidad del sistema educativo. Esta importancia se ha reflejado en las diferentes leyes educativas propuestas en España desde 1990 con la Ley Orgánica de Ordenación General del Sistema Educativo de España (LOGSE), donde ya se indicaba “La actividad evaluadora es fundamental para analizar en qué medida los distintos elementos del sistema educativo están contribuyendo a la consecución de los objetivos previamente establecidos. Por ello ha de extenderse a la actividad educativa en todos sus niveles, alcanzando a todos los sectores que en ella participan” (p. 3). Por su parte la Ley Orgánica 9/1995 (LOPEGCE), decía “La mejora de la calidad de la enseñanza exige ampliar los límites de la evaluación para que pueda ser aplicada de modo efectivo al

conjunto del sistema educativo” (p. 2). Más actualmente, la Ley Orgánica de Educación (LOE, 2006) establece “La evaluación se ha convertido en un valioso instrumento de seguimiento y de valoración de los resultados obtenidos y de mejora de los procesos que permiten obtenerlos. Por ese motivo, resulta imprescindible establecer procedimientos de evaluación de los distintos ámbitos y agentes de la actividad educativa, alumnado, profesorado, centros, currículo, Administraciones, y comprometer a las autoridades correspondientes a rendir cuentas de la situación existente y el desarrollo experimentado en materia de educación” (p. 4).

La evaluación, tal y como se concibe actualmente, está en el centro neurálgico del proceso de enseñanza y aprendizaje, donde el resto de procesos inmersos dentro del contexto educativo se ven influidos directamente. Esto hace que el concepto de evaluación educativa se encuentre definido por la influencia de una buena cantidad de factores diferentes, que de manera conjunta suman en su fundamentación. Esta diversidad de factores lleva a entender diferentes perspectivas de la evaluación educativa en función de los factores o circunstancias donde se contextualice (Casanova, 1995; Nevo, 1983). La situación educativa actual hace necesario diversificar las formas convencionales de evaluación que se utilizan actualmente para restablecer el enfoque en el proceso de aprendizaje. Hacer que la evaluación sea fundamental para el aprendizaje de los estudiantes se trata de volver a lo que enfatiza la investigación actual sobre evaluación educativa: el principal beneficiario de la evaluación debe ser el estudiante. Para lograr esto en el contexto de la evaluación, los estudiantes y tutores deben participar en un proceso de diálogo y retroalimentación (Hatzipanagos & Rochon, 2014). Las evaluaciones educativas deben representar plenamente las competencias que exige un mundo cada vez más complejo y cambiante. Las mejores evaluaciones pueden acelerar la adquisición de estas competencias si guían las acciones de los profesores y permiten a los alumnos medir sus progresos. Así, las tareas y actividades deben ser modelos dignos de la atención y la energía de profesores y alumnos (Pellegrino, 2020; Wilson, 2018).

1.2.- Conceptualización de Evaluación Educativa

Intentar hacer una aproximación al concepto de evaluación educativa no es una labor fácil. La evaluación es en concepto polisémico no uniforme, compuesto por factores y perspectivas que pudiera tomar diferentes posturas o planteamientos atendiendo al

contexto o la función que se pretende. La primera evaluación formal educativa se realizó en Estados Unidos en los años 1888-1889 por J. Rices y desde entonces, el concepto de evaluación ha ido evolucionando con práctica en función de su aplicación y el objetivo de la misma. Ralph Tyler, considerado como el padre de la evaluación educativa, conceptualiza la evaluación como un proceso esencial en la práctica educativa, mediante el cual se puede determinar en qué medida los objetivos educativos alcanzados a través de currículo educativos. Propone una sistematización del proceso mediante ocho fases: i. Establecer objetivos; ii. Ordenación de objetivos según su clasificación; iii. Definición operacional de los objetivos; iv. Establecer el mejor contexto para demostrar el logro de los objetivos; v. Explicar los objetivos; vi. Selección o desarrollo de medidas técnicas; vii. Recopilación de datos; viii. Comparación de los datos con los objetivos (Tyler, 1950). Desde estas primeras concepciones de la evaluación hasta nuestros días, el concepto de evaluación en educación ha sufrido una evolución notable. Uno de los factores que más ha contribuido a que la evaluación tenga un papel destacado en el ámbito educativo es que sea considerada como una disciplina científica, por parte de los profesionales de la educación, que tiene como objetivo la motivación y la regulación del aprendizaje (House, 1993). Esta nueva perspectiva a la hora de concebir la evaluación supuso una importante transformación a nivel conceptual pero aún más a nivel aplicado, llevándonos hacia la “cultura de la evaluación”, la cual ha trascendido más allá del ámbito educativo.

Tabla 1

Evolución historia de la conceptualización evaluativa

	Periodos	Conceptualización
1º	Hasta los años 20	Medida. Enfoque conductista
2º	Años 1930-40	Consecución de objetivos conocidos
3º	Hasta finales de los 60	Evaluación de la totalidad del sistema educativo
4º	Década de los 70 (España: LGE)	Valoración del cambio ocurrido en el alumno Valoración de programas y métodos
5º	Década de los 80	Evaluación cuantitativa/cualitativa

6º	Década de los 90 (España: LOGSE)	Formativa/diferenciada/integradora
7º	Comienzos S. XII (España: LOE)	Valoración del grado de adquisición de las competencias básicas. Demostración evidenciable de los resultados

En la Tabla 1, tratamos de hacer un resumen de la evolución histórica del concepto de evaluación. Este cambio conceptual podría concretarse en siete periodos o etapas diferentes. El primero de los periodos, se sitúa a finales del siglo XIX y principios del XX. En este periodo su concepción está íntimamente ligada a la psicología Conductista (Skinner y Watson) centrada principalmente en el propio acto de medir para obtener información. El segundo periodo está ligado a la figura de Tyler y a la concepción de la evaluación como un proceso sistemático donde se trata de evaluar la congruencia que existe entre los objetivos planteados y los logros alcanzados. El tercero de los periodos se desarrolló en la década de los sesenta y setenta en los Estados Unidos. Durante este tiempo la evaluación pasa a tener presencia en todos los aspectos del contexto educativo con el objetivo de aportar información para la mejora. Dos de los autores más influyentes fueron Lee Cronbach y Michael Scriven. Para Cronbach la evaluación era un proceso basado en criterios establecidos, el cual permitía obtener información que ayude a la toma de decisiones (Cronbach, 1963). Por su parte, Scriven (1966) fue quien propuso la separación conceptual de la evaluación sumativa y la evaluación formativa. El cuarto periodo está relacionado con las nuevas tendencias y enfoques asociados a la evaluación educativa que aparecen en la década de los 70. En este periodo se hace especial énfasis a la operativización de los objetivos a alcanzar, ya que estos determinarán el éxito que los programas educativos plantean alcanzar. Para ayudar a la operativización de los objetivos surge la evaluación criterial (*evaluación normativa*), que trata de ofrecer información real y concisa a los alumnos sobre los objetivos educativos planteado. El quinto periodo corresponde a los años 70. Este periodo está caracterizado por la aparición de los modelos de evaluación asociados a dos grades paradigmas, el cuantitativo y el cualitativo. Es en este periodo en el que en España aparece la Ley General de Educación, y con ella se introduce el concepto y la finalidad de la evaluación en educación. Ese mismo año, aparece el concepto evaluación continua (como heredera de la evaluación formativa propuesta por Scriven) (Rodríguez Diéguez, 2001), y desde ese momento quedó integrada

en los procesos de enseñanza y aprendizaje, y la evaluación formativa tomó una presencia prioritaria en todo acto educativo tratando de poner en paralelo el proceso de evaluación y la formación completa del alumno (García, 1989). El sexto periodo tiene lugar a partir de la década de los 90. En España se presenta la LOGSE (1990) la cual muestra una nueva concepción de evaluación. Diferencia entre la evaluación continua y global en etapas educativas de Educación Primaria, y una evaluación continua e integradora en Educación Secundaria. Con la concepción educativa de la evaluación que esta Ley propone, la evaluación se consolida en el centro de los procesos educativos, con la finalidad de aportar información útil para la orientación, la regulación y la mejora de los procesos educativos. Así la LOGSE propone que la evaluación sea sobre todo formativa, centrada en evaluar todo el proceso educativo, en todos los espacios del sistema educativo. Finalmente, el séptimo periodo aparece a comienzos del siglo XXI. En España coincide con la presentación de la LOE (2006), la cual hereda gran parte de la concepción que la LOGSE había planteado sobre la evaluación. Además, esta Ley establece que la evaluación debe de valorar en qué medida se alcanzan las ocho competencias educativas básica que esta misma Ley propone (comunicación lingüística; matemáticas; conocimiento e interacción con el mundo físico; digital; social-ciudadana; cultural y artística; de aprender a aprender; y autonomía e iniciativa personal).

Tal y como se ha mostrado la concepción de la evaluación educativa ha mostrado una naturaleza cambiante a lo largo de la historia. Este cambio se ha mostrado flexible y adaptativo a las características y necesidades educativa propias de la etapa (Bennett, 2015). En las últimas décadas, el uso más común de las pruebas educativas ha sido con fines institucionales. La magnitud de la evaluación llevada a cabo para servir a esos propósitos institucionales ha sido amplia, afectando a personas de todo el sistema educativo y, en parte por ese hecho, muy controvertida (Ravitch, 2013). Mientras que las pruebas con fines institucionales pueden no disminuir en términos absolutos, hay razones para creer que disminuirán en términos relativos a medida que la evaluación con fines de aprendizaje individual sea más frecuente (Gordon Commission on the Future of Assessment in Education, 2013). La creciente importancia de la evaluación con este fin está impulsada por muchos factores, como la necesidad de mejorar el rendimiento de los estudiantes, la creencia de que la información de la evaluación puede contribuir directamente al proceso de enseñanza y aprendizaje, la aparición de entornos de aprendizaje electrónicos, los avances en nuestra comprensión de la estructura de los

dominios y los caminos que los estudiantes probablemente tomen hacia el dominio de los mismos, y los avances en la medición y la ciencia de los datos (Bennett, 2015; Coates, 2020).

1.3.- Características de la Evaluación Educativa

En el apartado previo hemos podido ver que el concepto de evaluación educativa ha ido cambiando, tomando diferentes planteamientos y posturas a lo largo del tiempo y de las diferentes circunstancias. Se podría decir que todas estas concepciones de evaluación son correctas en el momento y en las circunstancias en las que fueron planteadas, pero en base a la concepción más actual de la evaluación, toda concepción evaluativa debe de mostrar una estructura básica que necesariamente debe de estar presente. Esta estructura tiene dos pilares. El primero de todos es concebir la evaluación como proceso susceptible a los cambios, abierto, y capaz de adaptarse a los diferentes contextos y situaciones educativas. Esta idea aleja la evaluación de la concepción puntual y aislada. La segunda idea principal emana de la primera, y postula que todo proceso evaluativo debe de cumplir unas fases o pasos enlazados que aseguren la sistematización del proceso. Estas fases son: 1ª Obtención de la información, a través de esta fase y haciendo uso de procedimientos válidos y fiables, podremos obtener diferentes tipos de datos relevantes en función de los objetivos de evaluación que nos permitan con seguridad confiar en los resultados de la evaluación; 2º Juicios sobre la información, en función de los datos obtenidos, de su análisis y valoración se formularán juicios sobre los objetivos planteados y las información recabada; y 3º Toma de decisiones, en esta fase final, en función de las valoraciones adoptadas sobre la información obtenida se tomarán decisiones que busquen la mejora del proceso educativo. Aunque esta es la estructura básica que toda evaluación debe de contener, dependiendo de diferentes factores pueden cambiar sus circunstancias como el momento de la evaluación (cuando llevar a cabo la evaluación), las funciones que la evaluación pretende (para qué evaluamos), los contenidos de la evaluación (qué evaluamos), el procedimiento que implementamos en la evaluación (cómo evaluamos), y los agentes evaluadores (quién evalúa).

De manera similar a la heterogeneidad vista sobre la conceptualización de la evaluación educativa, tratar de establecer las características esenciales de la evaluación

nos llevará de nuevo a diferentes posturas en función de diferentes autores. Por ejemplo, para Roger (1989) toda evaluación educativa debe ser integrada en el currículum, debe de ser formativa, ha de ser un proceso continuo, debe de ser un recurso recurrente, ha de plantearse en función de criterios establecidos, debe de permitir tomar decisiones, debe de permitir la cooperación entre agentes educativos. Por su parte, a lo propuesto por Roger (1989), Tindal y Martson (1990) hacen diferentes apreciaciones en función de las evidencias provenientes de la investigación. Estos autores apuntan a que las evaluaciones educativas deben de ser tanto generales como específicas, además apuntan a la necesidad de adoptar múltiples estrategias de evaluación, pero sobre todo enfatizan en la importancia y necesidad de una base empírica para la práctica evaluativa. Cardona (1994) también incide en la importancia de que la evaluación tenga un carácter científico tanto en lo que respecta a los instrumentos como a los procedimientos empleados, y además añade otras características: integral, comprensiva, indirecta, de referencia y cooperativa. En la actualidad, independientemente de las posturas que se adopten a la hora de concretar las características atribuidas a la evaluación educativa, en todos los casos esta evaluación debe de tratar de ser un proceso intencional y sistemático, integrado en el proceso educativo con la intención de orientar, y dar sentido a las diferentes acciones que suceden dentro de este proceso.

La evaluación educativa, al formar parte del propio proceso educativo, adoptará diversas funciones dependiendo de la necesidad que cada momento del proceso exija, por lo que las funciones de la evaluación siempre van a estar alineadas con los objetivos que se pretenden con la misma (Casanova. 1992). Dentro de las funciones más comúnmente atribuidas a la evaluación educativa podremos encontrar (Cardona, 1994): la función diagnóstica, la cual habitualmente suele realizarse en etapas educativas iniciales con el objetivo de conocer el punto de partida para iniciar una acción pedagógica; función regular, que permite obtener información sobre los procesos de aprendizaje con el fin de adaptar a la necesidades individuales; función previsor, mediante la cual nos permite adelantarnos y estimar posible logros para plantear acciones asociadas al cambios en el aprendizaje; función retroalimentadora, la cual desde una perspectiva formadora reconduce las acciones para adaptar los elementos de los modelos didácticos; y la función del control, la cual en ocasiones viene asociada a las demandas que la administración educativa propone para el control del logro dentro de las diferentes titulaciones.

Finalmente, la evaluación educativa puede adoptar una gran cantidad de modalidades en función de una gran cantidad de criterios, lo cual de nuevo hace que no sea fácil establecer una categorización universal para las modalidades de evaluación educativa. En función de establecer un criterio concreto, las modalidades de evaluación podrán ser unas u otras. Algunos autores han etiquetado a esta gran heterogeneidad asociada a la conceptualización, funciones y modalidades de la evaluación educativa como la gran “*maraña terminológica de la evaluación educativa*” (Lázaro, 1990). Según Casanova (1995), algunos de estos criterios de clasificación podrían ser en función de las fases o momento del proceso (inicial, continua y final), en función de la finalidad o de las funciones que la evaluación pretende (diagnóstica, formativa, sumativa), en función de la extensión del objetivo de la evaluación (global y parcial), según la procedencia de los agentes que llevan a cabo la evaluación (interna y externa), según los agentes que realizan la evaluación (auto-evaluación, hetero-evaluación y co-evaluación) y finalmente según el normotipo que se plantea en la evaluación (normativa o criterial).

El presente trabajo de tesis que aquí se presenta, tiene como objetivo centrarse en una de las modalidades de evaluación educativa. Concretamente esta tesis se centrará en analizar la autoevaluación como prueba evaluativa dentro del contexto académico, por lo que a continuación pasaremos a centrar el foco de atención en este tipo concreto de modalidad de evaluación.

Capítulo 2 AUTOEVALUACIÓN

2.1.-Introducción a la Autoevaluación

Como acabamos de indicar al final del apartado previo, de entre todas las modalidades de evaluación educativas, en esta tesis doctoral nos vamos a centrar en la Autoevaluación (AE en adelante). En el siguiente bloque trataremos de analizar el concepto de AE, y las diferentes perspectivas con el objetivo de mostrar la complejidad teórica de esta estrategia evaluativa, y simplificar su uso a través de recomendaciones educativas adecuadas. Haremos un repaso sobre el estudio de las diferentes tipologías de autoevaluación recogidas en la literatura y sus diferentes aplicaciones o repercusión en la práctica educativa, analizaremos los principales hallazgos empíricos asociados a esta técnica evaluativa y su implicación en la práctica docente.

Aunque más adelante definiremos más en detalle esta modalidad evaluativa, brevemente, nos referimos a AE cuando la persona que realiza la evaluación del aprendizaje es la misma persona que experimenta dicho aprendizaje, así en el contexto educativo AE es aquella evaluación que alumno hará sobre su propio aprendizaje. Las características de este tipo de evaluación hacen que el alumno sea un elemento activo dentro del proceso de evaluación y del mismo modo hace que el profesor pase a un segundo plano, y justamente son estas las características principales para que la AE haya atraído la atención, de manera positiva como negativa, tanto de docentes como investigadores educativos en las últimas décadas. La AE de manera general implica una amplia variedad de mecanismos y técnicas mediante las cuales los estudiantes describen y en la medida de lo posible asignan logros o fallos a las cualidades de sus propios procesos y productos de aprendizaje. Estos mecanismos implican que los alumnos sean capaces de monitorizar retrospectivamente los resultados de aprendizaje previos, y a su vez que sean capaces de informar sobre el nivel y la calidad del logro alcanzado (Baars et al., 2014; Raaijmakers et al., 2019).

Distinguiremos entre la AE como estrategia de aprendizaje que puede ser utilizada por los alumnos durante el proceso de aprendizaje, de AE como estrategia pedagógica a través de la cual los profesores piden a sus alumnos que estimen en logro en su trabajo. La distinción entre estos dos planteamientos de la AE nos ayudará a comprender las diferencias en la aproximación teórica a cada planteamiento para tener una visión comparada de ambos. La AE es un componente fundamental de la de la evaluación formativa ya que, como afirma Sadler (1989), es el propio alumno el que tiene que “*cerrar la brecha*” entre el rendimiento actual (revelado por la evaluación) y el estándar deseado. La AE se utiliza para abarcar los dos elementos clave en cualquier decisión de evaluación, por un lado, la identificación de criterios o estándares para ser aplicado sobre el propio trabajo, y por otro lado el poder formular juicios sobre la medida en la que el trabajo que hemos desarrollado se adapta a los criterios establecidos (Bound & Falchikov, 1989).

Desde la aparición de los trabajos publicados por Bound y Falchikov en los años ochenta (Bound & Falchikov, 1989; Flachivov & Bound, 1989), la popularidad de AE ha aumentado considerablemente, haciendo que sea un tema recurrente en trabajos de investigación sobre la evaluación en contexto educativo. La investigación en AE ha ido más allá del ámbito de la educación y de la psicología de la educación, llegando a áreas como psicología social (Dunning et al., 2004) o educación médica (Eva & Regehr, 2005).

En la práctica la AE también ha tomado una gran presencia dentro de las tareas docentes, una prueba de su importancia es que el uso de este tipo de técnica evaluativa es considerado un criterio de calidad dentro del sistema de evaluación de la calidad de la actividad docente universitaria (véase programa DOCENTIA de La Agencia Nacional de Evaluación de la Calidad y Acreditación, ANECA, 2021)

De entre los principales motivos que llevan a los profesores a promover la AE en los alumnos podemos destacar dos, por un lado, una motivación principalmente educativa y por otro lado la propia conveniencia del profesor. En primer lugar, uno de los principales deseos por parte de los de los profesores ha sido hacer que los alumnos se responsabilicen más de su propio aprendizaje mediante la participación activa en los propios procesos de aprendizaje: la evaluación de la propia competencia. En segundo lugar, existe una necesidad práctica de desarrollar procedimientos de evaluación que hagan un uso más eficaz de los recursos, utilizando más a los alumnos y menos a los de los profesores en las actividades de evaluación o, al menos, redirigiendo el esfuerzo de los profesores de calificar a planificar y moderar las actividades de evaluación. Este último motivo se combina a menudo con elementos del primero (Boud y Holmes, 1981). También se han tenido en cuenta otras consideraciones, como el deseo de los alumnos de tomar responsabilidad en los mecanismos de control educativo, los procedimientos de evaluación. Pero antes de comenzar a analizar las posibles implicaciones prácticas de la AE es necesaria hacer una aproximación conceptual a este tipo de evaluación educativa.

2.2.- Conceptualización de la Autoevaluación

Antes de conceptualizar la AE, es preciso realizar dos aproximaciones teóricas relacionadas con la concepción de la AE. Estas dos aproximaciones teóricas están relacionadas con las diferentes estrategias aplicadas al uso de AE: como estrategia del profesor y como estrategia del alumno. En el caso de la perspectiva de AE como estrategia del profesor, esta técnica de evaluación es entendida como un proceso institucional usado por el profesor como recurso educativo. Este uso de la AE es comúnmente utilizado por los profesores, con diferentes objetivos, a lo largo de las diferentes etapas educativas (Noonam & Duncan, 2005; Panadero et al., 2013). Mediante este uso, AE se convierte en parte del proceso académico y es considerado parte de la evaluación formativa (Black & Wiliam, 1998). Si consideramos la estrategia de AE desde el punto de vista del alumno,

esta perspectiva estará asociada a las teorías del Aprendizaje Autorregulado (AAR en adelante). Aunque hablaremos más adelante sobre AAR adelantando brevemente el concepto. La autorregulación es el control que las personas ejercen sobre sus pensamiento, acciones, emociones y motivaciones mediante estrategias personales para lograr los objetivos que se plantea (Zimmerman, 2000).

La situación educativa ideal del uso de AE desde un punto de vista pedagógico, sería aquella donde el profesor plantee el uso de esta técnica educativa desde su doble perspectiva, es decir, además de poder obtener información sobre la evaluación del logro del alumno, este es consciente del proceso y este proceso sirve de guía al alumno en la tarea de aprender como evaluar sus propios logros, con la intención de que este aprendizaje repercuta en una mejora de su autorregulación (Panadero & Alonso-Tapia, 2013). Este planteamiento busca que el uso de AE sea siempre con la máxima implicación o propósito pedagógico, aunque como ya veremos este planteamiento aplicado de la AE no siempre es el más frecuente ni en la literatura, ni en la práctica. Es importante concienciar sobre el uso de AE desde esta perspectiva más pedagógica, y de esta manera evitar que los profesionales de la educación que usan AE como técnica evaluativa puedan tener dudas sobre su utilidad pedagógica o sobre la precisión que los alumnos pudieran tener sobre sus estimaciones (Noonan & Duncan, 2005; Panadero et al., 2013; Vasileiadou & Karadimitriou, 2021).

Brown y Harris (2013), indicaban sobre la definición de AE que tanto los descriptores Tesouro en la base de datos de ERIC (Educational Research Information Center [U.S.], 2001), como el Tesouro de Psychological Index (Tuleya, 2007), asumen que la AE (*self-assessment*) es sinónimo de auto-valoración (*self-appraisal*), y que ambos términos se encuentran clasificados bajo la temática autoevaluación (*self-evaluation*). El Tesouro de Descriptores ERIC define autoevaluación como "la evaluación que los individuos hacen de sí mismos" (Educational Research Information Center [U.S.], 2001).

Una de las definiciones más clásicas sobre AE la describe como "actividad del alumno a través de la cual emite un juicio sobre su propio aprendizaje, específicamente sobre sus logros y sus resultados" (Bound & Falchikov, 1989, p. 529). Esta definición, aunque clásica, es una definición generalizada que marcó una delimitación clara en la concepción de AE. No obstante, esta definición deja algunos aspectos importantes a tener en cuenta

que se han ido señalando con el paso del tiempo sobre todo a través de las evidencias científicas.

Por su parte, Panadero, (2011, p. 78) introducen un concepto muy importante en la conceptualización de la AE “la autoevaluación es una evaluación cualitativa del proceso de aprendizaje, y de su producto final, realizada en base a unos criterios preestablecidos”. Esta definición introduce la idea de que toda estimación hecha por los alumnos sobre sus posibles logros debe de hacerse en base a unos criterios establecidos previamente conocidos por los agentes que intervienen en el proceso de aprendizaje. Además, esta definición al indicar el carácter cualitativo trata de alejar esta evaluación del único propósito de dar una puntuación numérica, y acercar el concepto a la idea de estimar el nivel o grado de logro en el aprendizaje. Aunque este nivel de logro podría ser expresado en términos numéricos, el número en si debe de contener más información que la numérica propiamente dicha. AE es un proceso reflexivo basado en criterios preestablecidos, de ahí que esta definición haga hincapié en la necesidad de la existencia y conocimiento de unos criterios preestablecidos para la evaluación. Esta idea, ha sido una de las principales ideas que más polémica y crítica ha suscitado dentro de la investigación sobre el uso de la AE en contextos educativos. Algunos estudios teóricos han tratado de alejar el concepto de AE de prácticas más asociadas a la calificación numérica aislada puramente sumativa etiquetándolas con otros términos para visualizar una diferencia con el término AE, como auto-graduación (del inglés self-grading), auto-calificación (del inglés self-marking) (Andrade, 2019; Andrade & Valcheva, 2009).

Otro planteamiento criticado sobre AE es entenderla como una evaluación enfocada al producto final del proceso de aprendizaje. Algunos autores critican esta concepción y la tachan de reduccionista aludiendo a que, si la AE tiene un carácter formativo y forma parte del proceso de autorregulación de los alumnos, el alumno va adquiriendo aprendizajes a lo largo de todo el proceso de aprendizaje, por lo que relegar el uso y sus posibles beneficios a el final del proceso podría estar limitando la potencialidad de la AE (Boekaerts & Cascallar, 2006; Greene & Azevedo, 2007; Zimmerman & Moylan, 2009). Para usar como referencia una concepción actual y amplia del AE tomaremos la que recientemente se ha propuesto en una de las revisiones más determinantes sobre AE hecha por Panadero, Brown, y Strijbos, (2016. p2.). En este trabajo los autores definen la AE como, “... gran variedad de mecanismos y técnicas a través de los cuales los estudiantes

describen (es decir, evalúan) y posiblemente asignan mérito o valor a (es decir, valoran) las cualidades de sus propios procesos y productos de aprendizaje”.

Como ya hemos comentado, la AE tiene una importante trayectoria en el ámbito práctico y en la investigación (Boud y Falchikov 1989; Leach 2012; Myyry et al., 2021; Tan 2012a), esto ha llevado a que exista una gran heterogeneidad en la manera de poner en prácticas la AE, haciendo que exista una gran amplitud y generalidad en la conceptualización del AE. Adicionalmente en la práctica, en ocasiones se hacen pocas distinciones entre diferentes maneras, mecanismos, o técnicas de AE (Panadero et al., 2016). Así, Panadero et al., (2016), en un intento de esclarecer la ambigüedad creada por la gran heterogeneidad en la concepción de AE, realizaron una revisión de todos los trabajos que se habían propuesto hacer un estudio topológico (sobre las tipologías) de la AE. Como resultados los autores encontraron 5 trabajos que analizaban la topología de la AE (cronológicamente: Boud y Brew, 1995; Tan, 2001; Taras, 2010; Alonso-Tapia y Panadero, 2010; Brown y Harris, 2013). En primer lugar, Boud y Brew 1995, presentaron una organización del AE basada en el interés de conocimiento que se pretendía con la prueba. En base a este criterio los autores proponían tres formatos diferentes de AE: en el que el interés se centraba en comprobar qué habilidades, conocimiento o ideas habían sido correctamente adquiridos (interés técnico), aquellos donde el interés se centraba en la comunicación un la interpretación de los elementos objeto de evaluación (interés comunicativo), y aquellos donde se busque que el alumnos sea capaz de comprometerse con los criterios de evaluación y juzgar su propio trabajo (interés emancipador). Estos autores además proponen distinguir entre AE y otros tipos de evaluaciones relacionadas con las auto evaluaciones, como auto-prueba (del inglés *self-testing*), auto-calificación (del inglés *self-rating*), o simplemente preguntas reflexivas. Para estos autores este tipo de formatos de evaluación no deberían de ser considerados AE ya que este tipo de evaluaciones no consiguen que los alumnos se comprometan con las normas y los criterios empleados en la evaluación. En segundo lugar, tenemos la tipología propuesta por Tan (2001). Este autor propone una tipología en base a la participación del profesor, relacionado con el propósito de la evaluación. Siguiendo este criterio en un continuo de mayor a menos participación del profesor, el autor propuso seis tipos diferentes de AE: auto-consciencia (*self-awareness*); auto-valoración (*self-appraisal*); evaluación auto-determinada (*self-determined assessment*); práctica de autoevaluación (*self-assessment practice*); tarea de autoevaluación (*self-assessment task*); y auto-calificación o

autocomprobación (*self-grading/self-testing*). La tercera clasificación de los tipos de AE fue propuesta por Taras (2010). Este autor propuso una clasificación basada en la idea del equilibrio de poder entre los estudiantes y sus profesores añadiendo el concepto de la transparencia en el formato de la AE. De más débil a más fuerte propuso: auto-calificación (*self-marking*); autoevaluación estándar (*sound standard of self-assessment*); modelo estándar (*standard model*); autoevaluación con retroalimentación integrada del tutor y de los compañeros; y diseño de contrato de aprendizaje. En cuarto lugar, encontramos la clasificación hecha por Panadero y Alonso-Tapias. Los autores desarrollaron una clasificación comprobada empíricamente donde proponían tres tipologías en función de la presencia y formato de los criterios de evaluación (Alonso-Tapia y Panadero, 2010; Panadero, Alonso-Tapia et al., 2013). Los tres tipos propuestos fueron: autoevaluación estándar; uso de rúbrica; y uso de guiones (*scripts*). Finalmente, Brown y Harris (2013) propusieron una clasificación de la AE en función del formato en el que esta se desarrolla. Los autores proponen tres tipos de AE: auto-valoración (*self-ratings*); auto-calificación (*self-marking*) o auto-estimación del rendimiento; y evaluación basada en criterios. En general, las cinco tipologías reflejan diferentes concepciones de la AE basadas en criterios tales como de poder, transparencia, uso y presencia de criterios de evaluación, implicaciones y procesos psicológicos, participación de los alumnos, finalidad de la autoevaluación, métodos de evaluación y/o apoyo pedagógico. Del estudio de la topología de la AE parece claro tener presente que la AE no es un constructo unidimensional. Puede estar mediado por una buena cantidad de variables o factores y estos pueden influir de manera importante en la naturaleza de la AE. En su revisión, Panadero et al., (2016) encontraron veinte categorías diferentes de implementación de la AE, desde concepciones más simples donde se piden a los alumnos que otorguen una calificación a un trabajo, sin más información, a otras formas más complejas de AE asociadas a análisis de logro basados en habilidades metacognitivas y retroalimentación formativa basada en criterios explícitos.

Tabla 2.

Taxonomía sobre AE propuesta por Andrade (2019)

Proceso		Producto	
Criterio		Criterio	
Si	No	Si	No

Formativo	Valoraciones de autoeficacia específicas de la tarea	Evaluación de los avances hacia objetivos específicos	Semáforo Control de la comprensión Autocontrol; Metacognición Diario de reflexión	Autoevaluación con rúbrica o lista de control Autovaloración	Crítica abierta del propio trabajo o comprensión
Sumativo	Juicios posteriores a la tarea sobre la capacidad basada en el rendimiento		Evaluaciones posteriores a la tarea sobre la eficacia de los procedimientos	Autocalificación	Autocalificación

Nota. Adaptado de Andrade (2019)

Más recientemente, tomando como referencia el trabajo sobre la clasificación de la AE propuesto en Panadero et al., (2016), Andrade (2019) ha planteado una nueva visión y organización a la hora de clasificar los tipos de AE. En su “*critical review*”, Andrade (2019), muestra una gran cantidad de definiciones de AE provenientes de diferentes ámbitos para indicar que, tomando en conjunto todas ellas podríamos definir cualquier tipo de evaluación. Tras esto, la autora indica que, desde su punto de vista, toda definición de AE podría ser correcta en función del propósito de la misma, pero lamentablemente esto es algo que no suele aparecer junto a la concepción de AE, y en muchas ocasiones tampoco cuando esta técnica se pone en práctica. Para Andrade (2010; 2019), el propósito de la AE es la retroalimentación, y tiene como objetivo informar a los alumnos sobre el ajuste de sus procesos y sus productos durante el aprendizaje. Lógicamente, esta concepción de la AE orientada al aprendizaje implica que debe ser una evaluación formativa. No obstante, tal y como plantea la autora, aunque la AE en su concepción formativa es el planteamiento más coherente desde el punto de vista pedagógico, ignorar otros tipos de AE con un propósito sumativo sería ignorar parte de una realidad evaluativa educativa. Siguiendo este planteamiento, Andrade (2019) propone una clasificación de las diferentes concepciones de AE en función del propósito formativo/sumativo y en función de si esta evaluación se centra en competencias, el proceso y/o el producto. La Tabla 2, presenta la taxonomía sobre AE propuesta por Andrade (2019). Mediante la taxonomía propuesta, la autora pretende dar respuesta a *qué* (competencias, proceso o producto), a *el por qué* (formativo o sumativo), y al *cómo* (métodos, incluyendo si incluyen o no estándares, por ejemplo, criterios) sobre la AE.

2.3.-Investigación actual sobre la Autoevaluación

En este bloque haremos un repaso sobre las principales temáticas de investigación sobre AE, así como los principales resultados encontrados dentro de cada una de las líneas de trabajo.

2.3.1.-Percepción del alumnado

Uno de los intereses que la AE ha suscitado dentro de la investigación educativa, es la percepción que los estudiantes tienen sobre esta prueba evaluativa. A este respecto, la percepción que los alumnos tienen sobre los diferentes factores o agentes involucrado en los procesos educativos siempre es de gran interés ya que el alumno es el principal agente dentro de estos procesos, por lo que conocer la percepción que este tiene sobre estos factores puede ayudar a mejorar su implicación y con ellos la repercusión sobre el aprendizaje.

Algunos de los estudios destinados a evaluar la percepción que el alumnado tienen sobre la AE han mostrado que estos parecen disfrutar cuando participan en los procesos evaluativos mediante la AE (Brookhart et al., 2004; McDonald, 2009; Papanthymou & Darra, 2019; Ross, Rolheiser et al., 2002). Concretamente los alumnos evalúan de manera especialmente positiva las situaciones de evaluación mediante AE cuando la experiencia les ayuda a asimilar y comprender mejor los criterios de evaluación establecidos, o en las situaciones en las que los alumnos deben de trabajar para alcanzar sus propios objetivos (McDevitt et al., 2008). Los estudios realizados han mostrado que los estudiantes valoran positivamente cuando los criterios de evaluación en una situación de AE se presentan mediante el uso de rúbricas. Los estudiantes también han manifestado que estas situaciones son especialmente útiles para comprender de manera más satisfactoria los criterios de evaluación establecido (Andrade, 2000; Andrade y Valcheva, 2009). No es sorprendente que las percepciones positivas de la AE fueran desarrolladas típicamente por estudiantes que participaron activamente en el tipo más formativo, por ejemplo, desarrollando sus propios criterios para una respuesta de autoevaluación efectiva (Bourke, 2014), o usando una rúbrica o lista de verificación para guiar sus evaluaciones y luego revisar su trabajo (Huang y Gui, 2015; Wang, 2017). Investigaciones anteriores sugirieron que las actitudes de los niños hacia la AE pueden volverse negativas si es

sumativa (Ross et al., 1998). Sin embargo, los estudiantes adultos informaron que incluso la AE sumativa era útil para ayudarlos a ser más críticos con su propia escritura y la de los demás a lo largo del curso y en los cursos posteriores (van Helvoort, 2012).

A pesar de que hay un interés demostrado de los alumnos por la AE, también podemos encontrar en la literatura algunos estudios en los que los alumnos muestran ciertas dudas sobre el uso de la AE. Podemos encontrar ocasiones en las que los alumnos no encuentran el sentido al uso de la AE, ni son capaces de entender el sentido pedagógico que pudiera tener, atribuyendo a AE connotaciones negativas, como una liberación de la responsabilidad del profesor. McDonald (2002) en su estudio encontró que la percepción que los alumnos tenía sobre la AE, era algo descontextualizada. Los estudiantes percibían como primera impresión que la AE se encontraba más cercana a la idea de habilidades de estudio autónomo, más que a las habilidades de reflexión sobre los méritos alcanzado con su propio trabajo. No obstante, estos mismos estudiantes cambiaron de parecer respecto a la concepción de AE con el tiempo, mostrando interés en este tipo de evaluación considerándola fuente de motivación y de utilidad para los procesos de aprendizaje (McDonald, 2009).

La falta de motivación o una desconexión del objetivo principal de la AE por parte de los alumnos puede llevar a que los alumnos se limiten a responder a la prueba rellenando lo que se le pide, sin realizar una reflexión retrospectiva sobre el aprendizaje alcanzado (Brookhart et al., 2004; Moss & Brookhart, 2019). Siguiendo en esta línea sobre la percepción negativa que los alumnos pueden llegar a tener sobre la AE, hay estudios que han mostrado que los estudiantes pueden llegar a mostrar dudas sobre si la AE se considera realmente una evaluación educativa (Brown et al., 2009; Brown, Peterson et al., 2009; Harris et al., 2009; Peterson y Irving, 2008), incluso estudios donde cuestionan su valor, informatividad y su posible efecto beneficioso (Brown et al., 2015; LaVoie y Hodapp, 1987), reclamando que el profesor no haga uso de esta técnica y en su lugar usar evaluaciones individualizadas del profesor (Andrews, 2016; Lasonen, 1995). En diferentes estudios, los estudiantes han manifestado tener una visión sobre la AE como algo aburrido, una apropiación inapropiada de la responsabilidad del profesor, y/o una manera de hacer trampas para poder obtener una evaluación diferente a la que obtendrían con una evaluación estándar (Gao, 2009; Harris & Brown, 2010; Johnson & Winterbottom, 2011; Peterson & Irving, 2008; Ross et al., 1998). Ross et al., (1998) descubrieron que los profesores hacían poco por explorar las ideas erróneas y las

preocupaciones de los alumnos sobre la autoevaluación, lo que hacía que muchos alumnos se volvieran cada vez más negativos hacia su uso.

Otro aspecto analizado sobre la percepción de la AE en alumnos está relacionado con el afrontamiento que los alumnos hacen sobre sus propias AE. Los estudios muestran que los alumnos no son partidarios e incluso muestran preocupación si creen que los resultados de su AE se harán públicos a sus compañeros, padres o profesores (Cowie, 2009; Harris & Brown, 2010; Raider-Roth, 2005; Ross et al., 1998, 2002). También encontramos estudios donde se informa de una actitud negativa hacia la AE si los alumnos saben que los resultados de esta se pondrán en común en el aula (Brown et al., 2015; Kasanen & Rätty, 2002). Esto resultados indican que los alumnos podrían de manera consciente generar sesgos en sus AE por temor a la impresión que el resto pueda tener sobre las valoraciones hechas. Esto sesgos pueden generar evaluaciones por debajo de lo esperado por miedo a que los alumnos sean considerados por lo demás como ególatras (Brooks, 2002). Por contrario, estos sesgos pueden hacer que los alumnos den AE elevadas para evitar que ser avergonzados frente a la clase (Harris & Brown, 2010).

Los estudios sobre la percepción de AE realizada con niños de edades más tempranas sugieren que tienden a tener una comprensión poco sofisticada de sus propósitos (Harris y Brown, 2013; Bourke, 2016) que podrían conducir a una implementación superficial de procesos relacionados. Por el contrario, estudios similares realizados con estudiantes de secundaria y universitario mostraron que estos estudiantes si entendían las funciones atribuidas a la AE (Ratminingsih et al., 2018), y que además de entenderlas, las consideraban una orientación útil para comprender la evaluación (Micán y Medina, 2017), así como para conseguir involucrarlos de manera responsable en sus procesos de aprendizaje (Lopez y Kossack, 2007; Bourke, 2014; Ndoye, 2017) haciendo que estos alumnos desarrollen habilidades de pensamiento profundo y crítico (van Helvoort, 2012; Siow, 2015), así como fomentando el aprendizaje autorregulado con el objetivo de establecer metas, planificar acciones, autocontrolar las acciones en los proceso de aprendizaje y reflexionar a lo largo del proceso (Wang, 2017).

Los resultados de estos estudios dan muestra de que los alumnos tienen reacciones muy diferentes respecto a la relevancia de la AE en el contexto educativo (Cowie, 2009; Harris et al., 2009). Por lo tanto, para que la AE sea una práctica eficaz en el aula, hay

que tener en cuenta las preocupaciones válidas que tienen los estudiantes sobre su legitimidad y práctica.

2.3.2.-Aprendizaje Auto-Regulado

En párrafos previos se ha hablado de las propiedades formativas atribuidas a la AE y de algunos de sus beneficios en los procesos de enseñanza y aprendizaje. Una buena parte de la investigación relacionada con la AE se ha centrado en analizar estos supuestos beneficios educativos. Estos beneficios están relacionados con atribuir a la AE una mejora de la interiorización del aprendizaje y del uso de estrategias de aprendizaje autorregulado por parte de los alumnos. Esta idea sugiere que la AE contribuye a la mejora de los procesos de aprendizaje de los estudiantes. Estas mejoras en procesos se relacionan por ejemplo con una mayor claridad de los objetivos de aprendizaje marcados, un aumento de la implicación de los estudiantes en el seguimiento de los procesos de aprendizaje y un ambiente propicio para la reflexión sobre los resultados de aprendizaje que los objetivos del proceso pretenden (Brown y Harris, 2013; Newton & Nation, 2020; Nicol y McFarlane-Dick, 2006; Panadero y Alonso-Tapia, 2013)

Antes de continuar analizando los principales hallazgos de la investigación donde relacionan la AE y el Aprendizaje Auto-Regulado (AAR en adelante), detengamos a hacer una breve conceptualización sobre AAR. Una de las definiciones más utilizadas es la propuesta por Zimmerman (2000, p. 14) la cual indique que AAR es “pensamientos, sentimientos y acciones autogenerados, planificados y adaptados cíclicamente a la consecución de objetivos personales”. El AAR se produce cuando los alumnos establecen metas, realizan un seguimiento, y manejan sus pensamientos, sentimientos y acciones para alcanzar esas metas.

El AAR es hoy en día una de las teorías educativas más populares a la hora de explicar el rendimiento académico de los estudiantes en contextos educativos. La popularidad de esta teoría se debe a que en su formulación incluye un gran número de variables relacionadas con el aprendizaje, como por ejemplo la orientación hacia las metas u objetivos, las estrategias específicas para las tareas a desarrollar, las estrategias metacognitivas asociadas a las tareas, la teoría de la atribución, etc. (para más detalle véase la revisión de Panadero, 2017). Como ejemplo para visualizar la importancia que

esta teoría está suponiendo en la investigación educativa enfocada en el rendimiento académico, Richardson et al., (2012) realizaron un meta-análisis enfocado a conocer la implicación de los 11 componentes atribuidos a la teoría de AAR sobre el rendimiento académico. Este estudio mostró que el uso de estrategias de AAR mostró ser un importante predictor significativo del rendimiento académico. Resultados similares también fueron informados años después por Broadbent y Poon (2015). El AAR está entre moderada y altamente correlacionado con un aumento del rendimiento académico (Zimmerman y Schunk, 2011).

En la literatura podemos encontrar diferentes modelos teóricos que tratan de mostrar cómo se presenta el AAR, y aunque podemos encontrar diferencias entre algunos de los planteamientos teóricos de cada uno de los modelos, la mayoría de ellos proponen una fase preparatoria, una fase de desempeño y una fase de evaluación, y cada una de estas fases consta de diferentes subprocesos (para más detalle véase Panadero, 2017). Uno de los más citados y usados sobre AAR es el propuesto por Zimmerman (2000). Este modelo se plantea como un modelo de AAR cíclico que se desarrolla en tres fases: previsión, desempeño y autorreflexión.

La AE y las estrategias de AAR se han relacionado teóricamente desde las concepciones más tempranas de la evaluación formativa (Black & Wiliam, 1998; Darling-Hammond et al., 2020). Como ejemplo de esta relación teórica, si tomamos el modelo de AAR propuesto por Zimmerman (2000), dos de los subprocesos de este modelo (*self-monitoring* y *self-evaluation*) tienen una clara similitud con la AE, ya que están orientados a la valoración del propio desempeño por parte del alumno. De manera similar Paris y Paris (2001) sugieren que la AE es crucial para la internalización de los estándares de evaluación, así gracias a esta los alumnos pueden de manera más efectiva asegurar una regulación del proceso de aprendizaje, siendo esto la primera fase de los modelos de AAR (previsión). Asumiendo estos planteamientos, la AE nos solo afecta a la autorreflexión, sino también a la previsión (por ejemplo, al proporcionar a los estudiantes criterios de evaluación, para que sean capaces de establecer objetivos realistas para la tarea), y a la ejecución (haciendo el seguimiento con más precisión al haber una comprensión más clara sobre el objetivo final que se pretende) (Andrade, 2010; Panadero & Alonso-Tapia, 2013).

Aunque ya hemos visto que existe una gran conexión entre la AE y AAR desde el ámbito teórico, es necesario hacer un repaso sobre qué es lo que se sabe de esta relación desde las evidencias empíricas que provienen de la investigación. Esta necesidad de afianzar esta relación en base a la evidencia ya ha sido demandada por diferentes autores. Por ejemplo, Topping (2003) en su trabajo concluye que, aunque las pruebas que sugieren que la AE afecta al AAR de los estudiantes son muy alentadoras, sería necesario más investigación para poder establecer una relación sólida. Por su parte, Brown y Harris (2013) en su revisión sobre AE apuntan a que las pruebas sobre la conexión entre AE y AAR no son sólidas, advirtiendo de que existen lagunas sobre qué estudiantes podrían verse beneficiados en la formación en AE.

Veamos a continuación algunos de los principales datos empíricos que apoyan esta relación. Así, Dignath et al., (2008), aportaba indicios sobre la importancia de integrar la planificación en los procesos relacionados con la AE. En su estudio meta-analítico revisaron las intervenciones basadas en AAR en educación primaria, y los resultados indicaron que las intervenciones metacognitivas basadas en la combinación de la planificación y seguimiento, o planificación y evaluación fueron las que fomentaron mayor uso de estrategias por parte de los estudiantes (con unos tamaños de efectos de $d=1.50$ y $d=1.46$, respectivamente), además de mostrar un efecto significativo sobre la motivación ($d=0.58$ y $d=1.59$), y más importante, sobre el rendimiento académico ($d=0.78$). Basándose en estos resultados, se ha sugerido que aquellas intervenciones que pretendan promover la AE deberían de llevarse a cabo antes de que los estudiantes comiencen a desarrollar las tareas. Estas intervenciones deberían de tener como objetivo, por ejemplo, que los estudiantes tengan acceso a los criterios de evaluación, así ellos pueden tener oportunidad de planificar, supervisar y evaluar su desempeño en cada una de las tareas propuesta en relación a los criterios de evaluación (Andrade y Valtcheva, 2009; Jonsson, 2014). Ramdass y Zimmerman, (2008) demostraron que el hacer que los estudiantes participaran en tareas de AE también hizo que aumentaras las habilidades relacionadas con la AAR. Se ha sugerido que la motivación, la autoeficacia, el compromiso, y el comportamiento de los estudiantes, así como la calidad de las relaciones entre estudiantes y profesor mejoran gracias a la participación de los estudiantes en tareas de AE (Allan et al., 2018; Glaser et al., 2010; Griffiths & Davies, 1993; Munns & Woodward, 2006; Olin & Sullivan, 2002; Schunk, 1996). Pero sin duda, una de las pruebas empíricas que evidencia de manera más clara la relación que existe en la AE y

las diferentes fases de los procesos de AAR la encontramos en Panadero y Romero (2014). En este estudio se facilitó a los alumnos criterios explícitos antes de realizar diferentes tareas que posteriormente serían evaluadas. Los resultados mostraron que facilitar los criterios a los alumnos tuvo un impacto positivo significativo, con respecto a alumnos que no tuvieron acceso a los criterios, en las fases de previsión, en la fase de ejecución y en la fase de autorreflexión. Estos resultados sugieren que las actividades relacionadas con la AE tuvieron una repercusión sobre todo en la activación de las fases de preparación y previsión.

En general, se podría decir que la investigación relacionada con AAR abarca tanto los procesos de aprendizaje, así como los productos de estos procesos, mientras que buena parte de la investigación basada en AE (y sobre todo la práctica de esta) está más centrada en juzgar el producto del aprendizaje de los alumnos y no tanto en el proceso. Este enfoque eminentemente práctico de la investigación sobre AE hace que esté por detrás en términos de desarrollo teórico respecto a la investigación sobre AAR que se basa en paradigmas experimentales de la psicología cognitiva (de Bruin y van Gog, 2012), mientras que la investigación sobre la AE está por delante en términos de implementación.

2.3.3.-Validez/Precisión/Consistencia

Como ya hemos visto, la AE entraña que los alumnos sean capaces de poder juzgar cuales han sido sus puntos positivos y sus puntos negativos en el proceso de aprendizaje. Esta estimación o juicio emitido por los alumnos permitirá al alumno saber qué salió mal durante el proceso con el fin de poder identificar sus errores para así no volver a cometerlo en el futuro y poder aprender de ellos (Dochy et al., 1999; Panadero & Brown, 2017). Es lógico pensar que un elemento clave en estas estimaciones es que estén lo más cercano posible a la calidad y/o cantidad de los logros de aprendizaje alcanzados por el alumno en el proceso que se evalúa (Kane 2006). Pero claro, en la AE estas inferencias o juicios hechos por los alumnos dependerán de la capacidad de los alumnos de emitir juicios razonablemente realistas o precisos (Butler 2011). Algunos autores han indicado que ese grado de realismo o verosimilitud en las estimaciones hechas mediante la AE, podría considerarse un tipo validez de constructo (Panadero, Romero et al., 2013). Para la psicometría (disciplina encargada de medir las variables psicológicas), la validez de

constructo es “el grado en que una prueba mide los significados que esta da” (Cronbach y Meehl, 1955). Aunque existe un amplio debate sobre la realidad objetiva de la AE (Tan 2012b), en educación este realismo es entendido mediante la alineación de la evaluación emitida por el estudiante (AE) con el rendimiento en la prueba o tarea administrada o juicios de expertos como profesores o incluso compañeros (Topping, 2003).

Esta característica sobre la AE es quizás una de las que más interés y popularidad han mostrado a lo largo del tiempo tanto por profesores como investigadores. Podemos encontrarla en la literatura con diferentes maneras de referirse a este concepto dependiendo del autor: validez, precisión, consistencia. Parte del interés que la investigación sobre AE ha mostrado en este aspecto específico, está relacionado con la preocupación que ha suscitado la falta de precisión, invalidez o inexactitud de la AE. La causa de esta preocupación es que este tipo de prueba comúnmente se aplica en contexto educativos que demandan de una evaluación para describir el progreso y los resultados de aprendizaje de los alumnos (Boud et al., 2013). Estas situaciones pueden llevar a los estudiantes a tomar decisiones sobre su aprendizaje en base a factores diferentes a sus capacidades, potencial o rendimiento observable (Brown y Harris 2013; Dunning et al., 2004).

Algunas de las principales fuentes que se han propuesto como posibles sesgos en los juicios sobre el aprendizaje son las mezclas de emociones fuertes con capacidad real (Baumeister et al., 2015) y sesgos cognitivos relacionados con defectos de memoria sobre lo aprendido (Finn y Metcalfe 2014). En la práctica docente, estamos acostumbrados a ver como los estudiantes toman decisiones basadas en concepciones erróneas de la realidad, por ejemplo no invertir más horas de estudio antes de una evaluación al pensar erróneamente que ya han invertido suficiente tiempo y que han adquirido los conocimientos necesarios (en base a errores metacognitivos), o decidir no interesarse en un tipo de conocimiento o en una opción profesional al pensar que no están preparados para ese tipo de conocimiento (Vancouver y Day 2005). En los ejemplos indicados, los alumnos están haciendo AE imprecisas sobre sus habilidades y/o logros que les llevan a tomar decisiones educativas poco acertadas. Uno de los objetivos de la práctica docente, y con ella del aprendizaje, es que a través de la retroalimentación los alumnos sean capaces de ajustar sus juicios y así corregir las AE irreales (Hattie y Timperley 2007). A través de la práctica, los alumnos serán capaces de adquirir aprendizajes y mecanismos

que les ayuden a calibrar los criterios y los estándares de valoradores externos con los que se evaluará su trabajo (Dochy et al., 1999).

Como ya hemos apuntado, este factor es uno de lo que más atención ha llamado sobre la investigación de la AE. Una buena parte de estas investigaciones se han centrado en evaluar la precisión de los alumnos en AE pidiéndoles a los estudiantes que estimen que puntuación o calificación obtendrían en un examen (Andrade, 2019; Boud y Falchikov, 1989; Dochy et al., 1999; Falchikov y Boud, 1989; Nelson y Dunlosky, 1991; Dunlosky y Nelson, 1994; Meeter y Nelson, 2003; Thiede y Dunlosky, 1999), reduciendo la AE a una mera puntuación o calificación. Aunque este tipo de AE son parte de la extensa tipología que podemos encontrar dentro de la AE, reducir la AE a la precisión de una puntuación aislada podría no ser tan importante como el contenido o las causas de dicha precisión (Andrade, 2019; Panadero et al., 2016; Ward et al., 2002). Desde el punto de vista pedagógico, tiene mucho más interés el poder conocer qué factores pueden influir, alterar o fomentar que los alumnos sean precisos a la hora de describir las cualidades de su trabajo (sus puntos fuertes o débiles susceptibles de mejorar). Conocer los factores, las situaciones, las partes del proceso que afecta a la precisión de la AE y conocer cómo puede afectar a la habilidad a AE y a su vez saber la transferencia positiva (los beneficios) que podría tener a los procesos de aprendizaje parece más complejo y a la vez interesante que el simple hecho de saber si los alumnos estiman una calificación cercana a la calificación que un experto/profesor le otorgaría en un prueba (Boud y Falchikov 1989). Así, el propósito más pedagógico, y que más debería de interesar a investigadores y sobre todo docentes de la precisión en AE, sería el relacionado con la información que los estudiantes puedan obtener a través del proceso de AE con el objeto de que les facilite el comprender las posibles vías de mejora en su proceso de aprendizaje.

Las revisiones sobre la precisión de la AE hechas por Ross (2006) y Brown y Harris (2013), apuntan que los estudiantes pueden llegar a ser consistentes en sus evaluaciones, pero al compararlas con otras medidas (experto) en ocasiones pueden llegar a ser inconsistentes. Las correlaciones entre AE y valoraciones de profesorado (v.gr., Sung et al., 2010), puntuaciones reales de exámenes estandarizados (v.gr., Luyten & Dolkar, 2010) o juicios usando rúbricas (v.gr., Sadler & Good, 2006) tendieron a ser positivas con valores que variaban entre .20 y .80, siendo pocos los estudios que mostraban correlaciones superiores a .60. No obstante, el autor sugiere que diferentes estrategias o entrenamiento hacen que los alumnos muestren más precisión en la AE.

A continuación, haremos un repaso de los factores que hasta el momento han sido propuesto por diferentes revisiones teóricas como posibles variables relacionadas con la variabilidad en la precisión que los estudiantes muestran en AE.

Edad y experiencia escolar

Los alumnos de edades más tempranas han mostrado ser más optimistas, indulgentes o generosos en sus AE que los alumnos mayores (v.gr., Kasanen et al., 2009; Ross, Rolheiser et al., 2002; Yan et al., 2020). No obstante, no está claro si esta falta de precisión se debe a la edad (desarrollo) o a la experiencia escolar. Stipek y Mac Iver (1989) en su revisión señaló que los niños en etapas tempranas utilizan criterios para juzgar sus logros intelectuales haciendo especial énfasis en el esfuerzo, el refuerzo social y el dominio, y con paso del tiempo estos criterios van madurando hacia una dependencia de la información más objetiva y normativa. Por su parte, Elder (2010) mostró que los alumnos de primer curso se centraban en características superfluas a diferencia de los alumnos de cuarto y quinto que emitían juicios más complejos, no obstante, todos los grupos se dejaban influir por las opiniones del profesor al tomar decisiones sobre sus logros o la calidad de sus trabajos. Se ha sugerido que las AE de los alumnos mayores además de ser inferiores que la de los alumnos más jóvenes, suelen mostrar una mayor precisión respecto a las valoraciones de expertos (v.gr., Bradshaw, 2001; Gascoine et al., 2017; Hewitt, 2005; Kaderavek et al., 2004).

Habilidad académica

Esta es una de los factores más estudiados relacionados con la precisión de la AE. Numerosos estudios (véase Brown y Harris, 2013) sugieren que la habilidad académica está relacionada con la precisión en AE, es decir, que los alumnos que muestran ser más competentes académicamente suelen ser más precisos a la hora de evaluar los logros académicos alcanzados. Además, se ha sugerido que los alumnos más competentes parecen ser más severos a la hora de evaluar sus trabajos mientras que los alumnos menos competentes suelen ser más indulgentes consigo mismo (v.gr., Laveault & Miles, 2002; Sung et al., 2010; Watt, 2000). Los resultados en estudios suelen apuntar a que los alumnos más inteligentes y capaces emiten AE que correlacionan más fuertemente con las evaluaciones del profesor que la de los alumnos menos competentes (v.gr., Brown et al., 2015; Laveault y Miles, 2002; Ng y Earl, 2008; Sung et al., 2010).

Dificultad de la tarea

Se ha indicado que la dificultad que la tarea pueda suponer para el alumno puede afectar a la precisión de los alumnos al AE (Bradshaw, 2001; Hewitt, 2005). Es intuitivo pensar que aquellas tareas más simples, familiares y predecibles para el alumno podrán probablemente hacer que estos puedan estimar de manera más acertada sus logros (Jones et al., 1977; Shapiro, 2014). Por el contrario, las tareas más difíciles requieren más atención y esfuerzo. Esta demanda podría influir o interactuar con los recursos cognitivos necesarios para controlar y evaluar el rendimiento. No obstante, estos podemos encontrar resultados con efectos contrapuestos en la relación de tareas fáciles asociadas a precisiones superiores de AE (Harris & Brown, 2018; Powel y Gray, 1995). En el caso de las tareas difíciles, la AE puede mejorar si se aporta retroalimentación sobre cómo se va progresando en la tarea (Lee y Gavine, 2003).

Base o criterio para la evaluación

Este factor también ha sido marcado como un factor clave y determinante no solo en la precisión de AE sino además de la influencia que esta tiene en el proceso formativo de los alumnos (Panadero et al., 2016). Algunos estudios muestran datos que indican que los alumnos valoran positivamente el uso de criterios basados en factores irrelevantes para medir la habilidad o el contenido que se pretende evaluar (como el esfuerzo empleado) (v.gr., Ross et al., 2002). No obstante, otro grupo de estudios han mostrado que las AE en las que los alumnos usan estándares o puntos de referencia específicos y concretos, en lugar de criterios subjetivos, están asociadas a una mayor precisión (Brown et al., 2015; Claes y Salame, 1975).

Otros factores relacionados con la precisión

Algunos estudios han sugerido que factores como el género, la cultura étnica y factores de personalidad de los alumnos puede ser un factor que influye en la precisión de la AE. Así, Blatchford (1997) descubrió en su estudio que, a lo largo del desarrollo de los alumnos, los estudiantes blancos (concretamente las chicas blancas) eran menos precisos en sus AE sobre su rendimiento académico que los estudiantes de etnia afroamericana. Por otro lado, Wells y Sweeney (1986) indicaron que los alumnos con una alta autoestima eran propensos a sobreestimar sus capacidades y logros mientras que

aquellos que tenían una autoestima más baja en ocasiones podían llegar a subestimar sus logros y sus habilidades (Brown et al., 2015).

También se ha sugerido que la formación en AE puede mejorar la precisión en AE. Así los resultados indican que aquellos alumnos que recibieron formación sobre el uso de una rúbrica para hacer las AE mostraron ser más precisos (Ramdass y Zimmerman, 2008). Algo similar se ha informado cuando se ha involucrado a los estudiantes en la creación conjunta de los criterios de evaluación mediante rúbrica y su posterior práctica (Ross, Rolheiser et al., 1998).

A modo de resumen, tal y como Panadero et al., (2016, p. 813) indica en relación a la validez/precisión de la AE “...parece tener propiedades psicométricas misteriosas”. Como ya hemos indicado previamente parece que AE es una parte integral de AAR y participa en la mejora de los resultados de aprendizaje (Ramdass y Zimmerman 2008). Por otro lado, los resultados de diferentes investigaciones han mostrado que asociados a la AE hay una buena cantidad de errores de medida e imprecisiones de validez desconocidos y esta falta de precisión o errores en ocasiones pueden llegar a eclipsar los beneficios formativos de AE (Brown y Harris 2013). No obstante, a pesar de que no hay una concreción exacta sobre la procedencia o las fuentes de imprecisión en esta medida, los beneficios de AE y sus aspectos positivos en el proceso de aprendizaje tienen pocas dudas. A pesar de eso, es necesario investigación que trate de esclarecer estas posibles fuentes de imprecisión en AE y aquellas variables o factores que pudieran afectar positivamente sobre la precisión de los alumnos en AE. Por otro lado, no existes evidencias de si AE más precisas tienen mayor efecto beneficioso sobre el aprendizaje que aquellas AE menos precisas, de ser así, conocer la manera de evitar imprecisiones en las AE nos llevarían a procesos de aprendizaje más beneficiosos para los alumnos. Que los alumnos tengan expectativas de autoeficacia positiva e ilusión positiva sobre su propio trabajo es un aspecto positivo que suele contribuir a la motivación de los alumnos en los procesos de aprendizaje, por el contrario, una ilusión negativa o pesimista sobre el trabajo realizado puede tener el efecto contrario (Butler, 2011). Por esta razón está claramente justificada la investigación centrada en comprender fuentes de imprecisión en la AE, así como factores asociados a una mejor precisión.

2.4.-Implicaciones de la AE para la práctica educativa

Pese a la popularidad que ha tomado la AE en los últimos años, la AE no siempre se aplica en las aulas. Por ejemplo, Hunter et al., (2006) encontraron que de los profesores de educación secundaria encuestados en Canadá (4.148), un 23% manifestaron no utilizar nunca la AE en sus aulas y el 58% indicaron que lo usaban lo mínimo posible. Algo similar se encontró en Finlandia, donde sólo la mitad de los profesores encuestados manifestaron usar la AE (Lasonen, 1995). Esta ausencia de la AE en la práctica docente es probablemente debida a la tensión que los profesores pudieran llegar a mostrar al pensar que los resultados de la evaluación están controlados por los alumnos y que esto podría afectar a un resultado académico irreal (Harris & Brown, 2010; Volante & Beckett, 2011).

En la literatura, se ha señalado la importancia de que los alumnos trabajen la habilidad de AE por cuatro razones fundamentalmente. La primera de las razones, como ya hemos esbozado previamente, es que los alumnos entrenados en AE habitualmente suelen mostrar una mejora en sus procesos de aprendizaje, además de una mejora en el rendimiento académico (Andrade, 2019; Brown y Harris, 2013; Panadero et al., 2012; Ramdass y Zimmerman, 2008). De manera adicional, habitualmente los alumnos que suelen ser más precisos respecto a sus estimaciones, suelen estar asociados a resultados académicos mayores, respecto a los alumnos menos precisos (Brown y Harris, 2013). En segundo lugar, entrenar a los alumnos en AE está relacionado con el aumento del uso de estrategias de aprendizaje autorregulado por parte de los alumnos (Kostons et al., 2012), de la misma manera que aquellos alumnos que muestran buenas habilidades de AAR suelen al mismo tiempo mostrar mejores estrategias de AE (Lan 1998). Las habilidades de AE ayudan a los estudiantes a regular su propio aprendizaje, ya que esta les exige una supervisión metacognitiva de sus procesos y su desempeño en relación los estándares marcados, las expectativas y los objetivos o metas marcadas (Panadero y Alonso-Tapia, 2013). En tercer lugar, el entrenamiento de las habilidades de AE pueden mejorar la autoeficacia ende los estudiantes a la hora de realizar las tareas que se les exigen (Ramdass y Zimmerman, 2008). En cuarto lugar, la AE está relacionada con el empoderamiento de los estudiantes en el proceso de evaluación, los convierte en parte activa en este proceso, lo cual los hace más responsables de su aprendizaje y del uso de

estrategias de autorregulación (Black & Wiliam, 1998; Nicol & McFarlane-Dick, 2006; Tan, 2012; Taras, 2010). En resumen, la AE parece desempeñar un papel importante en el éxito académico y, en consecuencia, existe una amplia defensa de la AE como un poderoso proceso de aprendizaje educativo. De hecho, se ha recomendado que se desarrolle un plan de estudios para la AE para que los estudiantes puedan desarrollar habilidades de AE en el contexto de las actividades del aula (Andrade, 2019; Brown y Harris 2014).

Capítulo 3: INVESTIGACIÓN EXPERIMENTAL

Este capítulo está destinado a presentar de manera resumida los aspectos más importantes de los cuatros trabajos que constituyen la parte experimental de la tesis doctoral. La información se presentará de manera resumida y accesible, para mayor grado de detalle los trabajos originales pueden consultarse en los Anexos, donde se encuentran integrados. Estos trabajos son (por orden de presentación):

León, S. P., Pantoja-Vallejo, A., & Nelson, J. B. (2021). Variability In The Accuracy Of Self-Assessments Among Low, Moderate, And High Performing Students In University Education. *Practical Assessment, Research, and Evaluation*, 26(1), 16. <https://scholarworks.umass.edu/pare/vol26/iss1/16>

León, S. P., Pantoja-Vallejo, A., & Ferrero, M. (2021, artículo enviado). Influence of class position on academic performance and Self-Assessment: an exploratory study from a Frequentist and a Bayesian perspective. *Studies in Higher Education*.

León, S. P., & García-Martínez, I. (2021, in press). Feedback and evaluative experience as decisive factors in student self-regulation. *Publicaciones*.

León, S. P., Augusto-Landa, J. M., & García-Martínez, I. (2021). Moderating Factors in University Students' Self-Evaluation for Sustainability. *Sustainability*, 13(8), 4199. <https://doi.org/10.3390/su13084199>

3.1.- Objetivos

General

- Analizar la precisión de la AE de los estudiantes universitarios, y explorar qué posibles factores podrían estar relacionados con variaciones en la previsión o validez de la AE, así como evaluar cuales podrían ser sus implicaciones en la práctica docente.

Específicos

- i. Evaluar en qué medida la competencia que los alumnos tienen con el contenido objeto de evaluación puede llegar a afectar a la precisión de los alumnos al evaluar sus logros mediante el uso de una prueba de AE. Al mismo tiempo, se han analizado las implicaciones que los posibles efectos sobre la validez de la prueba relacionada con este factor podrían tener sobre el aprendizaje y la práctica docente (León, Pantoja-Vallejo y Nelson, 2021).
- ii. Analizar cómo factores ambientales asociados a un mejor rendimiento académico, como la posición que los alumnos escogen para sentarse en clase, puede estar relacionada con la habilidad de los alumnos a la hora de evaluar su rendimiento académico a través de la AE. Adicionalmente se analizó la confianza que los alumnos tienen en sus juicios y su implicación para la práctica docente (León, Pantoja-Vallejo y Ferrero, 2021).
- iii. Evaluar el efecto que tiene, sobre la precisión de las AE y las evaluaciones a compañeros, la experiencia de los alumnos realizando evaluaciones. Adicionalmente, analizamos el posible efecto aditivo que pudiera tener sobre la precisión de las evaluaciones que los alumnos reciban retroalimentación sobre las evaluaciones emitidas por parte de los docentes (León y García-Martínez, 2021).
- iv. Analizar mediante un modelo basado en Ecuaciones Estructurales las relaciones que pudieran existir entre factores psicosociales/personalidad como el compromiso académico y las estrategias de estudios y la habilidad de los alumnos al emitir AE sobre su rendimiento académico (León, Augusto-Landa y García-Martínez, 2021).

3.2.- Metodología

3.2.1.- Muestra

La muestra utilizada en los cuatro trabajos que componen esta tesis doctoral está constituida por alumnos de los Grados de Educación (Educación Infantil, Educación Social, y Educación Primaria) de la universidad de Jaén. La suma total de la muestra de los cuatro trabajos ha sido de 905 estudiantes (León & García-Martínez, 2021, n= 71; León, Augusto-Landa & García-Martínez, 2021, n=630; León, Pantoja y Ferrero, 2021, n= 130; León, Pantoja y Nelson, 2021, n= 74). Las muestras de dos de los trabajos fueron grupos naturales de clase (León & García-Martínez, 2021; y León, Pantoja y Nelson, 2021). Todos los estudiantes que participaron en estos estudios lo hicieron voluntariamente. Las edades de los estudiantes que participaron en estos estudios fueron de $M = 22.08$ ($SD = 3.97$). La proporción de estudiantes respecto al sexo en los estudios fue entre 83.96-79.23% para mujeres y 16.03-20.76% para hombre, siendo estas proporciones una distribución acorde a la proporción entre hombres y mujeres en la población total de estudiantes de los grados de educación en España (Instituto Nacional de Estadística, 2020).

3.2.2.- Instrumentos

En todos los estudios se midió el rendimiento académico de los alumnos a través de diferentes herramientas o instrumentos. Mediante estos instrumentos de medida se evaluaban los diferentes logros académicos alcanzados en las diferentes asignaturas de los grados de educación. Todos los instrumentos de evaluación fueron exámenes de contenido de asignatura salvo en el caso de León y García-Martínez (2021), que fue mediante evaluaciones realizadas a los trabajos de prácticas desarrollados y presentado por grupos de clase. Todas las pruebas daban la oportunidad al alumno de conseguir una puntuación de 10 puntos, por lo que la escala de medida tanto para la prueba de evaluación de contenido, como para la de AE siempre fue de 0 a 10, siendo 0 la puntuación menos y 10 la mayor. Las pruebas de evaluación de León, Pantoja-Vallejo y Ferrero, (2021) y León, Augusto-Landa y García-Martínez, (2021) fueron exámenes de múltiple alternativa de respuesta con 20 y 24 preguntas (respectivamente) y 3 posibles alternativas, donde sólo una opción podía ser correcta. Todas las preguntas de cada una de las pruebas fueron

seleccionadas al azar dentro de un banco de preguntas creadas previamente por el profesor. Para analizar la fiabilidad de las pruebas de elección múltiple y debido a su carácter nominal se usó la prueba Lambda de Guttman (λ^2) la cual mostró una fiabilidad de .82. En estos exámenes para corregir la probabilidad de acertar gracias al azar se aplicó la siguiente fórmula de corrección:

$$Nota = Acierto - \frac{Errores}{k - 1}$$

k = número de alternativas de respuesta

En el caso de León, Pantoja-Vallejo y Nelson, (2021) el examen consistió en seis preguntas abiertas de desarrollo (dos de desarrollo largo y cuatro de desarrollo corto). Los alumnos podían obtener 3 puntos a través de cada una de las preguntas largas y 1 punto a través de cada una de las preguntas cortas. En el caso de León y García-Martínez, (2021) la prueba consistió en la evaluación de la presentación del trabajo de prácticas de la asignatura. La prueba evaluaba cada una de los apartados principales de este trabajo asignando un porcentaje de puntuación máximo para cada apartado. Para los estudios donde no se usó una prueba de opción múltiple (León, Pantoja-Vallejo y Nelson, 2021; y León y García-Martínez, 2021), la evaluación se realizó en base a una rúbrica que se facilitó a los alumnos previamente para que se familiarizara con ella (véase un ejemplo en Tabla 1 en León, Pantoja-Vallejo y Nelson, 2021). Estas rúbricas establecían una relación entre el nivel de logro y la descripción del logro esperado para ese nivel. Antes del uso de las rúbricas, la validez de estas fue evaluada por 5 profesionales de manera independiente, mostrando un coeficiente de correlación intraclase de .962.

Todos los exámenes a su comienzo tenían un texto introductorio en el que se explicaba las instrucciones del examen, los criterios de evaluación, así como el peso de cada una de las partes en la nota final. Juntos a las instrucciones del examen se presentaban las instrucciones para la prueba de AE. En estas se indicaba que una vez finalizado el examen podrían emitir su AE sobre el trabajo realizado en relación a los criterios establecidos. En el caso de León, Pantoja-Vallejo y Ferrero, (2021), se le pedía a los alumnos que indicasen en qué filas de clase se habían sentado preferentemente a lo largo de todo el curso, pudiendo escoger entre a) las tres filas delanteras, b) las tres filas intermedias, y c) las tres filas finales. Finalmente, en el caso de León, Pantoja-Vallejo y Ferrero, (2021) y León, Augusto-Landa y García-Martínez, (2021) los alumnos podrían emitir mediante una

escala de 5 puntos un juicio sobre la confianza que tenían sobre la AE emitida, siendo 1 “nada seguro” y 5 “completamente seguro”.

Finalmente, en León, Augusto-Landa y García-Martínez, (2021) se usaron dos escalas para medir dos factores: compromiso y estrategias de estudio. La escala utilizada para medir compromiso académico fue la versión en español de UWE-9 (Schaufeli et al., 2006). Esta escala a través de 9 ítem y una escala Likert de 7 puntos medían 3 sub-factores (vigor, dedicación y absorción). La Escala usada para medir estrategias de estudio fue una adaptación del Cuestionario de Técnicas de estudio (Herrera-Torres & Lorenzo-Quiles, 2009). Mediante 12 ítem y escala Likert de 7 puntos medían 3 sub-factores (estrategias de organización, estrategias previas al estudio y estrategias de estudio). Las escalas fueron implementadas en Google Form para su digitalización y fácil acceso a los alumnos.

3.2.3.- Procedimiento

En todos los estudios, los alumnos durante el curso recibieron al menos una sesión en las que el profesor de la asignatura les hablaba de las AE. Se les facilitaba los criterios de evaluación e incluso se programaban sesiones en las que los alumnos pudieran poner en práctica dichos criterios en situaciones simuladas por el profesor. Este acercamiento a los criterios a su uso se hizo con el objetivo de que los alumnos estuvieran familiarizados con los criterios y fueran capaces de utilizarlos en la práctica antes de la prueba.

Tanto en León, Pantoja-Vallejo y Nelson, (2021), León, Augusto-Landa y García-Martínez, (2021), como en León, Pantoja-Vallejo y Ferrero, (2021) la prueba de evaluación se realizó al final del curso mediante un examen. El día de la prueba, el profesor tras repartir los exámenes explicó las instrucciones de la prueba, así como las instrucciones para realizar la prueba de AE y el objetivo de la misma. Se indicó que la participación en el examen, o en la prueba de AE era totalmente voluntario y la realización o no de la prueba no afectaría a la nota final. Del mismo modo se informó que aquellas personas interesadas en conocer más sobre el objetivo del estudio podrían contactar con el profesor. Todos los procedimientos realizados en los estudios de esta tesis doctoral siguieron los estándares éticos que guían la investigación con personas, según la Declaración de Helsinki (AMM, 2013).

En el caso de León, Augusto-Landa y García-Martínez, (2021) a lo largo del curso, el profesor de la asignatura en una sesión de clase, pidió a los alumnos de manera voluntario y haciendo uso de los dispositivos móviles accedieran a un formulario de Google Form el cual el enlace de acceso se encontraba ubicado en el espacio virtual de la asignatura en la plataforma oficial de docencia de la Universidad de Jaén. Al acceder al cuestionario los alumnos se encontraban una primera página de consentimiento informado donde se presentaban las instrucciones y los objetivos del cuestionario. De estar conforme en participar el alumno debería de clicar en una casilla que decía “*He leído y entendido las condiciones del estudio y deseo participar en el mismo de forma voluntaria.*” y esto le daba paso a un bloque de preguntas generales y a dos bloques de preguntas tipo Likert, uno para la escala de compromiso y otro para la escala de estrategias de estudio.

Finalmente, y a diferencia del resto de trabajo León y García-Martínez, (2021) este trabajo se hizo de manera longitudinal a lo largo de las diferentes sesiones de prácticas. Para este estudio, la clase se dividió en dos grupos (FB y N-FB). A su vez, si hicieron subgrupos de no más de 4 alumnos los cuales deberían de trabajar de manera conjunta en el trabajo de prácticas de la asignatura. Una vez concluido el trabajo, en las semanas finales del cuatrimestre, los grupos de alumnos uno a uno, alumnos debían de presentar el trabajo en clase. Mientras tanto el resto de grupos debían de evaluar (haciendo uso de la rúbrica facilitada) los trabajos presentados por el resto de grupos y por su propio grupo. Esto haría que cada grupo emitiera una AE de su trabajo y una evaluación del trabajo del resto de grupos de clase. La única diferencia entre los dos grandes grupos de clase (FB vs N-FB) fue que en uno de los grupos (FB) una vez emitidas las evaluaciones por parte de los grupos, el profesor ofrecía feedback sobre las fortalezas y debilidades de los trabajos presentados, mientras que en el grupo N-FB esta información no se daba.

3.2.3.- Análisis de los datos

Para todos los estudios realizados se fijó un α de .05. Los análisis de los estudios León, Pantoja-Vallejo y Nelson, (2021) y León, Augusto-Landa y García-Martínez, (2021) fueron realizados con el software r, y los análisis de los estudios León y García-Martínez, (2021) y León, Pantoja-Vallejo y Ferrero, (2021) con el software libre jamovi (The jamovi Project).

Las variables dependientes mediadas en los estudios incluidos fueron, las autoevaluaciones hechas por los alumnos AE (SSA en los trabajos de *Student Self-Assessment*) en las diferentes pruebas, las evaluaciones hechas por los expertos en las pruebas (*Exam* en los trabajos), la habilidad o precisión en las AE obtenida al restar las puntuaciones de AE las puntuaciones obtenidas en Exam para cada alumno (*SSA skill* en los trabajos). Tanto AE como Exam se medirán en una escala de 0 a 10 donde 0 es la puntuación menor y 10 la mayor. La escala de SSA skill podrá ir de -10 a +10, las puntuaciones positivas indicarán sobreevaluación, las puntuaciones negativas subevaluación y las puntuaciones cercanas a 0 denotarán precisión en la AE. En dos de los estudios (León, Augusto-Landa y García-Martínez, 2021 y León, Pantoja-Vallejo y Ferrero, 2021) se midió los juicios de confianza de los alumnos sobre las AE (CJ en los trabajos).

En el estudio de León, Pantoja-Vallejo y Nelson, (2021), se analizó la precisión que los alumnos tenían sobre sus AE con respecto a las evaluaciones hechas por el profesor. Así en este estudio por un lado tendremos la evaluación del alumno (SSA) y por otro la evaluación del profesor (Exam). Tras analizar las relaciones entre SSA y Exam mediante el coeficiente de correlación de Pearson (r), se hicieron agrupaciones de los alumnos en función del logro alcanzado en el examen, creando los grupos: Suspenso (Fail) con notas en el examen inferiores a 5; Aprobado (Pass) con notas entre 5 y 7; Notable (Good) con notas entre 7 y 9; y Sobresaliente (Outstanding) con notas superiores a 9. Tras la agrupación se analizó la relación entre las SSA y Exam mediante la prueba T de Student. Finalmente, y dada la distribución anormal de los participantes entre los grupos y que los datos provenían de un grupo natural, se hizo una estimación poblacional basada en un análisis bootstrapping de 5.000 remuestreos (con remplazamiento) mediante el paquete de *r dabestr* (Ho et al., 2018).

Para el estudio León, Pantoja-Vallejo y Ferrero, (2021), como variable de agrupación se usó la posición que los alumnos escogieron para sentarse durante el curso (filas primeras, filas intermedias y filas finales). Dado que los antecedentes mostraron que la posición en clase podría llevar a resultados nulos al predecir el rendimiento académico, hicimos una aproximación analítica desde una doble perspectiva: Frecuencialista y Bayesiana. Así, en el caso de encontrar un resultado nulo, gracias a la perspectiva Bayesiana podríamos calcular la fuerza en las evidencias de las hipótesis establecidas

tanto la hipótesis del investigador como la hipótesis nula a través del Factor de Bayer (Navarro et al., 2019).

En León y García-Martínez, (2021), las variables analizadas por un lado fue la precisión que los grupos mostraron evaluando el rendimiento del resto de los grupos sesión a sesión (Group Assessment-Skill, GA-Skill en el trabajo) y sobre su propia evaluación (Group Self-Assessment skill, GSA-skill en el trabajo). Para analizar como la precisión de las evaluaciones fue evolucionando a lo largo de las 9 sesiones de clase se realizó un ANOVA de medidas repetidas para cada uno de los dos grupos (FB vs N-FB). Se midió el tamaño del efecto en el contraste mediante *generalized eta squared* (η^2_G).

En el caso de León, Augusto-Landa y García-Martínez, (2021), se incorporaron dos factores medidos mediante escala. Se analizó las propiedades psicométricas (validez y fiabilidad) de las escalas con un doble objetivo. Por un lado, analizar que estas escalas muestras unas propiedades psicométricas aceptables en la muestra escogida, y por otro lado con el objetivo de escalas las puntuaciones brutas obtenidas en las escalas por las cargas factoriales obtenidas mediante el Análisis Factorial Confirmatorio (AFC) y así poder tener una medida más sensible. Finalmente, con las puntuaciones escaladas de los factores (Compromiso y Estrategias de estudios) y las variables dependientes, se propuso un modelo basado en Ecuaciones estructurales de matrices de covarianzas (CB-SEM) con el objetivo de comprender las relaciones que se establecen entre las variables del modelo. Tanto AFC como CB-SEM se realizaron con el paquete de r *lavaan* (Rosseel, 2012).

Finalmente, con el objetivo de resumir los efectos encontrado en los trabajos que componen esta tesis doctoral, se realizó un meta-análisis del efecto de la precisión en la AE con respecto las puntuaciones del experto mostrado por los alumnos en las diferentes mediadas de las investigaciones realizadas en la tesis. Este análisis será el único que no aparece en los artículos que componen el trabajo ya que es un análisis propio de la tesis en su conjunto. Para ellos se calculó la puntuación media estandarizada de la diferencia de las puntuaciones entre AE y Exam aplicando la Ecuación 4.18 y 4.19 de Borenstein et al., (2009) incluyendo la corrección del factor J calculado con la Ecuación 4.22 y 4.23. Para calcular la varianza se usó la Ecuación 4.20 y 4.24 de Borenstein et al., (2009). Dado que algunos de los trabajos mostraban más de una variable dependiente o diferentes grupos se realizó un análisis multinivel de modelo de efecto aleatorio usando la función *rma.mv* del paquete *metafor* de r (Viechtbauer, 2010). Los tamaños de los efectos calculados se

presentarán mediante una g de Hedges, la cual mostrará la diferencia entre las puntuaciones manifestadas por los alumnos en los diferentes estudios mediante sus AE y las puntuaciones otorgadas por los profesores expertos. Tamaños de efecto positivos muestran un efecto general de sobreestimación de las AE sobre las valoraciones expertas y tamaños de efecto negativos muestran un efecto general de infravaloración. Tamaños cercanos a 0 indican ausencia de efecto sugiriendo una tendencia general a la precisión entre ambas puntuaciones.

3.2.- Resultados

En esta sección presentaremos de manera resumida y accesible los principales resultados encontrados en cada uno de los trabajos que componen esta tesis doctoral. En la medida de lo posible, trataremos de interpretar los resultados, evitando el uso de expresiones numérico-matemáticas. Para mayor grado de detalle por favor, véase los trabajos incluidos en el capítulo 4.

En el caso de León, Pantoja-Vallejo, y Nelson (2021), las puntuaciones de los Exámenes (Exam) y las puntuaciones de los alumnos en la prueba de AE (SSA) mostraron ser muy semejantes $M_{exam} = 6.62$ y $M_{ssa} = 6.75$. La correlación entre estas dos variables resultó ser significativa $r = .66$. El analizar las relaciones entre estas variables en función de las agrupaciones hechas según la nota del examen (suspensos, aprobados, notables y sobresalientes) los resultados mostraron que, los alumnos suspensos sobreestimaron su AE, indicando puntuaciones por encima de las obtenidas en el examen. Los alumnos del grupo de aprobados (el más numeroso) fueron los que mostraron mayor precisión. Y finalmente tanto los alumnos del grupo notable como los del grupo sobresaliente mostraron subestimar sus AE, estimando puntuaciones por debajo de las puntuaciones obtenidas en el examen. Con la intención de poder tener una medida de la precisión de los alumnos en su AE, calculamos la precisión de esta restando las puntuaciones dadas por los alumnos a las puntuaciones obtenidas en el examen y analizamos la precisión en función de los grupos. Así puntuaciones cercanas a 0 indicarán buena precisión, puntuaciones por encima de 0 indicarán sobreestimación y puntuaciones por debajo de 0 subestimación. Dado que el estudio se realizó con un grupo natural de clase, y que el número de participantes por grupo era desigual y en algunos grupos muy reducido, se

realizó un análisis Bootstrap para poder estimar a través del efecto muestral el efecto poblacional. Para ellos se calculó el tamaño del efecto de la precisión de cada uno de los grupos a través de la g de Hedges. Estos análisis nos sirvieron para comprobar que el efecto poblacional replicaba lo encontrado a través de nuestro efecto muestral, sobreestimación en alumnos suspensos, buen ajuste en los aprobados y subestimación en notables y sobresaliente.

Por su parte, en León, Pantoja-Vallejo, y Ferrero (2021), advertidos por la literatura y en previsión de un posible efecto nulo en alguna de nuestras manipulaciones, se analizaron los resultados desde una doble perspectiva, la Frecuencialista y la Bayesiana. Esto nos permitiría analizar la potencia de la probabilidad tanto de la hipótesis nula como de la hipótesis alternativa. De manera similar a lo visto en el estudio anterior, las puntuaciones promedio entre AE y el examen se mostraron cercanas ($M_{ssa} = 6.70$ y $M_{exam} = 7.01$) con una correlación de $r = .47$. El análisis de la posición en clase desde la perspectiva Frecuencialista mostró que esta variable no tuvo influencia sobre la nota que los alumnos obtuvieron en el examen. Por el contrario, sí mostró influir en las puntuaciones sobre AE y sobre los juicios de confianza que los alumnos mostraron sobre estas AE. Los alumnos que se sentaban en las filas cercanas al profesor (las primeras) mostraron mayores AE y tenían mayor confianza sobre estas AE que los alumnos que se sentaban en las filas más alejadas al profesor (las del final de la clase). Desde la perspectiva Bayesiana, los resultados sobre la influencia de la posición sobre las puntuaciones del examen mostraron moderadas/fuertes evidencias a favor de la hipótesis nula, la cual plantea que la posición en clase no influye en las puntuaciones del examen, reafirmando el efecto nulo encontrado. De manera similar a lo encontrado mediante los análisis Frecuencialista las evidencias fueron fuertes a favor de la hipótesis del investigador para las variables dependiente AE y juicio de confianza, es decir, que la posición de los alumnos mostró modular las puntuaciones tanto para AE como para CJ.

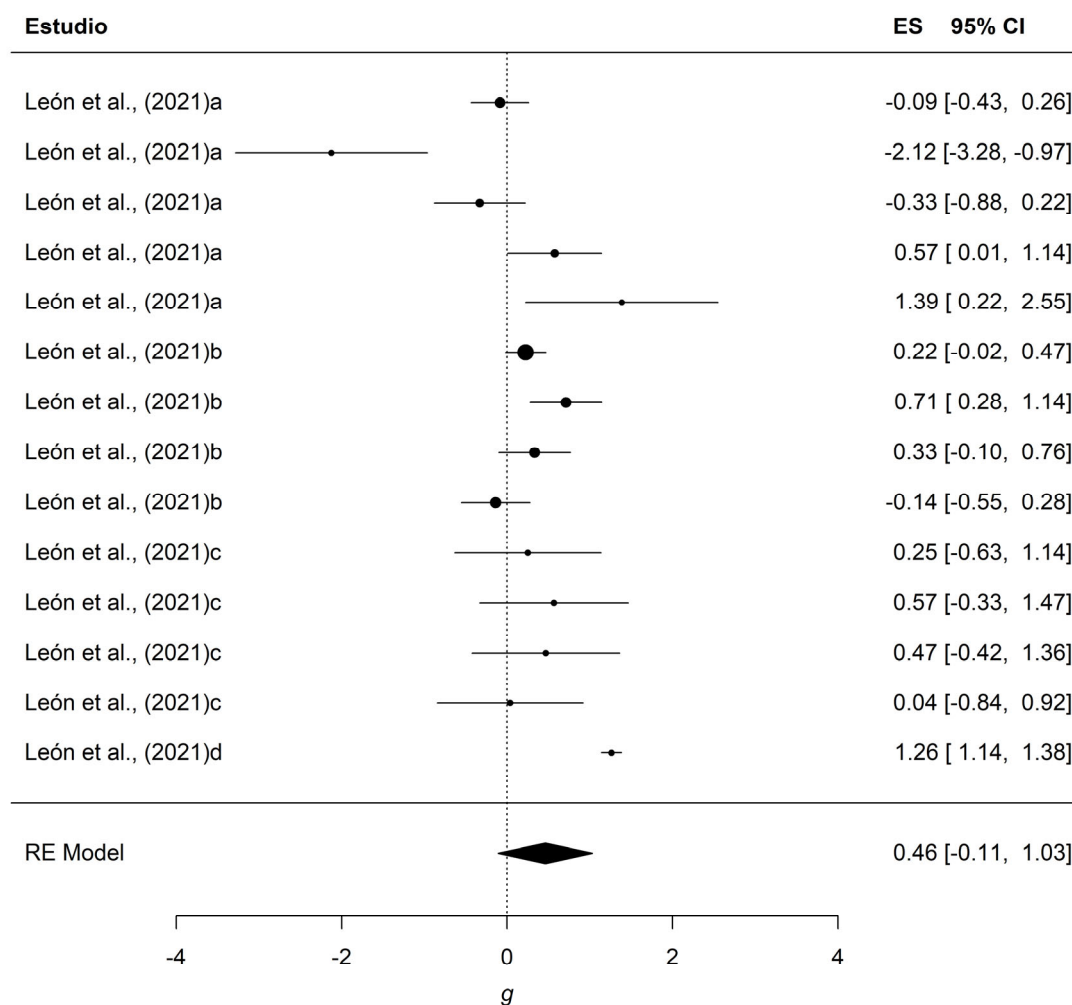
Los resultados en León y García-Martínez (2021) mostraron que todos los grupos de alumnos de prácticas mejoraron la precisión de las evaluaciones hechas sobre los trabajos de los compañeros a lo largo de las sesiones. No obstante, la mejora fue significativamente mayor y más precisa para los grupos que recibieron feedback del profesor una vez realizadas las evaluaciones de los grupos. Concretamente, los grupos de alumnos que recibieron feedback (FB) en las últimas sesiones de evaluación mostraron evaluaciones cercanas a 0 (indicando buena precisión) mientras que en el caso de los

alumnos que no recibieron feedback (N-FB) aunque la mejora se mostró de sesión a sesión, las últimas sesiones seguían mostrando sobreestimación de las evaluaciones grupales. Este efecto se mostró de igual manera en las AE que los grupos hicieron sobre su propio trabajo.

León, Augusto-Landa y García-Martínez (2021), propusieron un modelo teórico explicativo basado en Ecuaciones Estructurales en el que analizaban la relación entre las variables dependiente analizadas hasta ahora AE, Exam, CJ y Precisión en AE (SSA-Skill) y dos factores psicosociales medidos a través de escalas de autoinforme (compromiso académico y estrategias de estudio). Antes del analizar el modelo teórico propuesto, se analizó las propiedades psicométricas de las escalas usadas con el objetivo de comprobar la validez interna estructural y la fiabilidad en la muestra seleccionada mediante un Análisis Factorial Confirmatorio (AFC). Este análisis mostró una excelente validez y fiabilidad de las escalas y a continuación se pasó a analizar el modelo estructural propuesto. Los resultados del modelo mostraron un ajuste excelente, indicando que el modelo teórico explicativo predecía de manera excelente el modelo empírico que emanaba del conjunto de datos. Los resultados de las relaciones propuestas en el modelo mostraron que tanto el compromiso como las estrategias de estudio predecían significativamente la varianza mostrada en las puntuaciones del Examen y AE, así como la precisión en AE. Además, en el caso de las estrategias también predecía los juicios de confianza (CJ) que los alumnos mostraban sobre las AE.

Finalmente, y a modo de resumen de los resultados sobre la precisión de AE informados en los trabajos que componen esta tesis doctoral, realizamos un meta-análisis de los efectos encontrados. La Figura 1 muestran las estimaciones puntuales y los intervalos de confianza de todos los estudios incluidos en el meta-análisis, junto con su media meta-analítica. El efecto promediado a través de todos los efectos reportados mostró ser pequeño/intermedio $g = 0.46$, 95% CI [-0.11, 1.03], pero no resultó ser significativo $z = 1.59$, $p < .112$. El nivel de heterogeneidad fue amplio y significativo, $Q(13) = 174.04$, $p < .001$. Como se puede observar en la figura los valores de los efectos son muy diversos, muestra de ello es la falta de significación en el efecto. El efecto total positivo muestra una superioridad de las puntuaciones de AE sobre las puntuaciones del examen.

Figura 1. Forest plot de los trabajos incluidos en la tesis.



Nota. Las letras en subíndice de cada estudio corresponden con a= León, Pantoja-Vallejo, y Nelson (2021), b= León, Pantoja-Vallejo, y Ferrero (2021), c= León García-Martínez (2021), y d= León, Augusto-Landa y García-Martínez (2021).

3.2.- Conclusiones

En el siguiente bloque trataremos de presentar de manera resumida las principales conclusiones a las que podemos llegar a través de los resultados obtenidos en los estudios que componen esta Tesis.

El objetivo general que esta Tesis Doctoral se planteó fue analizar la validez o precisión de las AE de los estudiantes universitarios frente a las evaluaciones emitidas por expertos, así como posibles fuentes de imprecisión sobre esta validez. En términos generales si analizamos la precisión de las AE emitidas en los trabajos incluidos, podríamos decir que hay un cierto consenso entre las puntuaciones informadas por los alumnos en las AE con respecto de las aportadas por los profesores (Andrade, 2019, Brown y Harris, 2013, Panadero et al., 2016), aunque con una gran cantidad de variabilidad. La tendencia general, tal y como hemos podido observar al meta-analizar los resultados de los cuatro estudios, es a una sobreestimación ($g = 0.46$), es decir, los alumnos tienden a estimar sus logros por encima de lo evaluado por el experto, aunque como hemos podido observar debido a la variabilidad del efecto el resultado general resulta no ser significativo. Quedarnos con esta visión general del efecto estaría dejando de lado muchos matices que a lo largo de los estudios hemos podido analizar cómo han podido llegar a modular la precisión que los alumnos tienen sobre la AE. Teniendo en cuenta estas apreciaciones, considerar a la AE en contextos educativos como herramienta de evaluación educativa válida sería dejar a un lado parte de la realidad.

El primero de los objetivos específicos propuestos planteó analizar en qué medida la competencia que los alumnos tienen con el contenido de la asignatura puede ser una fuente de variabilidad en la precisión de los alumnos frente a la AE. La investigación llevada a cabo en León, Pantoja-Vallejo y Nelson (2021) se realizó con el propósito de dar respuesta a este objetivo. En la primera fase analítica de este estudio, cuando se tomó como unidad de análisis el grupo completo de clase, los resultados obtenidos mostraron una correlación aceptable entre las puntuaciones de AE y el examen. Estos resultados parecían corroborar la tendencia general sugerida en algunos estudios (Andrade, 2019, Brown y Harris, 2013), donde se muestra una aceptable precisión del AE emitida por los estudiantes. No obstante, un análisis más exhaustivo de los resultados, donde se agrupó a los alumnos en función de las competencias mostradas en la asignatura, mostró imprecisiones importantes. Esto nos lleva a pensar que la AE, en este caso concreto, podría ser válida como prueba evaluativa (mostrando un buen ajuste entre AE y la evaluación del profesor) tan solo para los alumnos aprobados, pero no para los alumnos de los extremos. Estos datos nos ponen en aviso sobre la idea de que la competencia en la materia podría ser un factor que afecte a la precisión de la AE (Brown y Harris, 2013, Panadero et al., 2016). Estas imprecisiones han sido asociadas en la literatura con sesgos

metacognitivos, concretamente con el efecto denominado *Dunning-Kruger* (Kruger & Dunning, 1999). Este efecto se ha descrito como un sesgo cognitivo en el que las personas evalúan erróneamente su capacidad cognitiva como mayor de lo que en realidad es. Los autores proponen que los estudiantes no sólo carecen de conocimientos sobre el contenido, sino que también carecen de habilidades metacognitivas para reconocer que no poseen esos conocimientos. El sesgo cognitivo, en los “estudiantes menos capaces”, se atribuye a una ilusión interna sobre sus propias capacidades cognitivas, lo que los lleva a sobreestimar su rendimiento. En el caso de los “estudiantes más capaces”, el sesgo cognitivo puede surgir de una percepción errónea acerca de que la evaluación externa de sus competencias es más rigurosa de lo que es, lo que los lleva a subestimar su rendimiento en la evaluación (Kolić-Vehovec et al., 2021). Tomando estos resultados en conjunto nos llevaría a concluir que usar la AE como prueba de evaluación educativa podría ser acertado para solo una parte de la población, pero generaría sesgos en los extremos de esta población. Aunque claro, los sesgos en la validez no son incompatibles con su capacidad formativa.

Continuando con el análisis de posibles factores que pudieran ser fuente de variabilidad o imprecisión en las AE de los estudiantes, se planteó el segundo objetivo específico. Este pretendía analizar como un factor ambiental asociado a la modulación del rendimiento académico, como es la posición que los alumnos eligen en clase durante las sesiones (Bergtold et al., 2019; Clément y Bukley, 2017), pudiera también modular la capacidad de los alumnos a lo hora de estimar sus logros mediante la AE. Este objetivo se analizó mediante el trabajo desarrollado en León, Pantoja-Vallejo y Ferrero (2021). En este estudio, y de manera similar a lo visto en el estudio previo, al tomar todas las puntuaciones de las clases, volvió a aparecer la tendencia general, mostrando una correlación aceptable entre AE y Exam. Pero, de manera similar a lo visto en León, Pantoja-Vallejo y Nelson (2021), al agrupar a los alumnos, atendiendo la posición en la que los alumnos se sentaron durante las sesiones de clase, de nuevo aparecieron sesgos de precisión en los alumnos sobre las AE. En este estudio, aunque la posición de los alumnos en clase mostró evidencias de no modular el rendimiento académico de los alumnos, si mostró ser una variable influyente en la precisión de los alumnos sobre sus AE y sobre la confianza que estos tenían sobre las AE. Previamente la literatura ya había asociado la posición de los alumnos en clase con una mejora en el rendimiento y tal y como hemos visto en León, Pantoja-Vallejo y León (2021), la competencia de los

alumnos parecía influir en la precisión, por lo que la relación entre estas variables podría ser lógica. De nuevo, el efecto de sobreevaluación en la AE mostrada por los alumnos de las primeras filas junto al efecto sobre el CJ podrían estar relacionados a los posibles sesgos metacognitivos de los alumnos que podrían correlacionar con factores motivacionales o personales.

Avanzando en el análisis de los posibles factores asociados a fuentes de variación en la precisión de AE se propuso el tercero de los objetivos específicos. Este objetivo planteó analizar en qué medida la experiencia que los alumnos tenían realizando evaluaciones podrían hacer mejorara la precisión de las AE de los alumnos. Adicionalmente se planteo evaluar si ofrecer retroalimentación tras las evaluaciones podría mejorar el posible efecto obtenido por medio de la experiencia. Este tercer objetico fue abordado en el estudio realizado por León y García-Martínez (2021). Los resultados de este trabajo parecen indicar que la práctica o la experiencia que los alumnos tienen a la hora de realizar evaluaciones puede ser otro factor capaz de afecte a la precisión de los alumnos a la hora de realizar evaluaciones de compañeros o las suyas propias (Sung et al., 2005). En nuestro estudio, ambos grupos fueron mejorando la precisión de las evaluaciones con el paso de las sesiones y con ellos con la práctica realizando evaluaciones. Además, los resultados de este estudio mostraron que ofrecer retroalimentación a los alumnos después de sus evaluaciones les permitía mejorar las futuras evaluaciones haciendo que este factor interactúe sobre los propios beneficios que la experiencia tiene sobre la precisión de la AE.

El cuarto y último objetivo planteado en esta Tesis Doctoral, propuso analizar cómo dos factores psicosociales ampliamente asociados al rendimiento académico (Quílez-Robres et al., 2021) como el compromiso académico y las estrategias de estudios, podrían predecir la precisión de los alumnos a la hora de estimar sus logros mediante la AE. Para analizar este último objeto León, Augusto-Landa y García-Martínez (2021) plantearon un modelo teórico explicativo analizado mediante ecuaciones estructurales (SEM). Los resultados de este estudio nos permitieron, además de tener una visión de la relación entre las variables dependiente analizadas en esta tesis (la AE, Examen y los juicios de confianza), nos permite analizar cómo diferentes variables psicosociales/personales asociadas directamente con el rendimiento académico, pueden predecir las puntuaciones de la variable dependiente medidas. Concretamente, Los resultados de este estudio han mostrado relaciones positivas entre los factores compromiso académico y estrategias de

estudios y las modalidades de evaluación, así como la capacidad de autoevaluación de los estudiantes y la precisión de los mismos. En concreto, este estudio demostró que estos factores predecían significativamente la varianza de las puntuaciones derivadas de las modalidades de evaluación (Exam y SSA), así como la capacidad de autoevaluación de los estudiantes (SSA-skil). Estos resultados son coherentes con investigaciones anteriores (Diseth, 2011; Munns, & Woodward, 2006). Los resultados encontrados pueden interpretarse entendiendo que los estudiantes con un alto compromiso académico y con buenas estrategias de estudio mostraron buenos resultados académicos y fueron precisos a la hora de evaluar sus logros, por lo que podríamos inferir que estos factores están asociados a altos niveles de rendimiento académico. En cuanto al análisis de los juicios de confianza sobre las evaluaciones (CJ), los resultados sugirieron que los estudiantes con mejores estrategias de estudio tendrán más confianza a la hora de evaluar su propio rendimiento (SSA). Sin embargo, esta confianza no se encuentra en los estudiantes con altas puntuaciones de compromiso. En este sentido, la literatura sugiere que la confianza en los juicios también puede estar relacionada con el desarrollo de habilidades metacognitivas vinculadas al AAR (Schuster et al., 2020).

En definitiva y a la vista de los resultados encontrados, parece que más allá de la precisión general (o promediada) que puede llegar a mostrar la AE en contextos educativos, hay factores que pueden llegar a modular esta relación entre el logro percibido por el alumno y la evaluación del profesor o profesional. Conocer qué factores perjudican esta imprecisión podría llevar a los docentes a tratar de controlarlos y así impedir su influencia negativa. Del mismo modo, conocer qué factores o situaciones facilitan la precisión nos ayudará a general un contexto favorable que ayude al alumnado a poder evaluar retrospectivamente su trabajo y determinar de manera precisa el logro alcanzado.

Como ya se ha señalado, el carácter formativo de la AE es algo probado, pero hasta donde nosotros conocemos, son pocas las investigaciones que tratan de analizar la relación entre la precisión de la AE y su capacidad formativa. Es decir, es cierto que hacer que los alumnos pasen por una situación de evaluación es beneficioso para el aprendizaje, pero hay pocas evidencias sobre si estos beneficios son mayores en el caso de que el alumnado sea más preciso (véase Panadero et al., 2017). La investigación relaciona precisión en AE con mayor rendimiento académico y mayor competencia en la materia

(León, Augusto-Landa, y García-Martínez, 2021 y León, Pantoja-Vallejo y Nelson, 2021), pero estas relaciones se analizan a posteriori, por lo que no se podría establecer una relación causal. Quizás esta laguna podría ser un futuro nicho de trabajo experimental para el futuro de la investigación en AE.

Bibliografía

- Agencia Nacional de Evaluación de la Calidad y Acreditación (2021). Programa de evaluación DOCENTIA. <http://www.aneca.es/Programas-de-evaluacion/Evaluacion-institucional/DOCENTIA>
- Ainscow, M. (1988). Beyond the eyes of the monster: an analysis of recent trends in assessment and recording. *Support for learning*, 3(3), 149-153. <https://psycnet.apa.org/doi/10.1111/j.1467-9604.1988.tb00088.x>
- Allan, F., Herbert, A., Peter, P., & Bridget, S. (2018). *Teachers investigate their work: An introduction to action research across the professions*. Routledge.
- Alonso-Tapia, J., & Panadero, E. (2010). Effect of self-assessment scripts on self-regulation and learning. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 33(3), 385–397. <https://doi.org/10.1174/021037010792215145>.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 1–7.
- Andrade, H. L. (2010). Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). Routledge.
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Front. Educ.* 4:87. <https://doi.org/10.3389/educ.2019.00087>
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice*, 48(1), 12-19. <https://doi.org/10.1080/00405840802577544>
- Andrews, C. D. (2016). *Student self-assessment: Teachers' definitions, reasons, and beliefs*. Brigham Young University.
- Arredondo, S. C. (2002). Didáctica de la evaluación: hacia una nueva cultura de la evaluación educativa. En Arredondo, S. C., y Botía, A. B. (Ed.), *Compromisos de la evaluación educativa* (pp. 1-33). Pearson Educación.
- Asociación Médica Mundial (2013). Declaración de Helsinki. 64ª Asamblea General, Fortaleza, Brasil, Octubre. *Principios Éticos para las Investigaciones Médicas en Seres Humanos; Asociación Médica Mundial: Ferney-Voltaire, France*.
- Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective

- monitoring of problem solving. *Learning & Instruction*, 33, 92–107.
<https://doi.org/10.1016/j.learninstruc.2014.04.004>.
- Baumeister, R. F., Alquist, J. L., & Vohs, K. D. (2015). Illusions of learning: irrelevant emotions inflate judgments of learning. *Journal of Behavioral Decision Making*, 28(2), 149–158. <https://doi.org/10.1002/bdm.1836>
- Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, 39(1), 370–407.
<https://doi.org/10.3102/0091732X14554179>
- Bergtold, J. S., Yeager, E. A., & Griffin, T. W. (2019). Spatial dynamics in the classroom: Does seating choice matter?. *PloS one*, 14(12), e0226953.
<https://doi.org/10.1371/journal.pone.0226953>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
<https://doi.org/10.1080/0969595980050102>
- Blatchford, P. (1997). Pupils' self assessments of academic attainment at 7, 11 and 16 years: Effects of sex and ethnic group. *British Journal of Educational Psychology*, 67(2), 169–184. <https://doi.org/10.1111/j.2044-8279.1997.tb01235.x>
- Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation?. *Educational Psychology Review*, 18(3), 199-210. <https://doi.org/10.1007/s10648-006-9013-4>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Boud, D. J., & Holmes, W. H. (1981). Self and peer marking in an undergraduate engineering course. *IEEE Transactions on Education*, 24(4), 267-274.
<https://doi.org/10.1109/TE.1981.4321508>
- Boud, D., & Brew, A. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education*, 18, 130–135.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*, 18(5), 529-549. <https://doi.org/10.1007/BF00138746>
- Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation In Higher Education*, 38(8), 941–956.
<https://doi.org/10.1080/02602938.2013.769198>.

- Bourke, R. (2014). Self-assessment in professional programmes within tertiary institutions. *Teaching in Higher Education, 19*, 908–918.
<https://doi.org/10.1080/13562517.2014.934353>
- Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge J. Educ. 46*, 97–111. <https://doi.org/10.1080/0305764X.2015.1015963>
- Bradshaw, B. K. (2001). Do students effectively monitor their comprehension? *Reading Horizons, 41*(3), 143–154.
https://scholarworks.wmich.edu/reading_horizons/vol41/iss3/2
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1e13.
<http://dx.doi.org/10.1016/j.iheduc.2015.04.007>.
- Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57*, 213–227.
<https://doi.org/10.1023/B:EDUC.0000049293.55249.d4>
- Brooks, V. (2002). *Assessment in secondary schools: The new teacher's guide to monitoring, assessment, recording, reporting and accountability*. Open University Press.
- Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks: Sage.
- Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: reframing selfassessment as a core competency. *Frontline Learning Research, 3*, 22–30. <https://doi.org/10.14786/flr.v2i1.24>.
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactiveinformal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning & Instruction, 19*(2), 97–111.
<https://doi.org/10.1016/j.learninstruc.2008.02.003>
- Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Self-regulatory beliefs about assessment predict mathematics achievement. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What*

students can tell us about assessment for learning (pp. 159–186). Charlotte, NC: Information Age Publishing.

- Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444-457. <https://doi.org/10.1080/0969594X.2014.996523>
- Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances: new results and future directions. *International Journal of Educational Research*, 50(4), 251–256. <https://doi.org/10.1016/j.ijer.2011.08.006>
- Cardona A., J. (1994). *Metodología innovadora de evaluación de centros educativos*. Sanz y Torres.
- Casanova, M. A. (1992). *La evaluación, garantía de calidad para el centro educativo:(manual para la evaluación interna de los centros docentes)*. Luis Vives.
- Casanova, M. A. (1995). *Manual de evaluación educativa*. Madrid, La Muralla.
- Claes, M., & Salame, R. (1975). Motivation toward accomplishment and the self-evaluation of performances in relation to school achievement. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*, 7(4), 397–410.
- Clément, G., & Bukley, A. (2017). Where do you sit in class? A study of spatial positioning during two courses of different duration. *J. Hum. Psychol*, 1, 1-7. <https://doi.org/10.14302/issn.2644-1101.jhp-17-1460>
- Coates, H. (2020). Reforming Assessment. In *Higher Education Design* (pp. 91-115). Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-15-9216-4_7
- Cowie, B. (2009). My teacher and my friends helped me learn: Student perceptions and experiences of classroom assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 85–105). Charlotte, NC: Information Age Publishing.
- Cronbach, L. J. (1963). Evaluation for course improvement. *Teachers College Record*, 64(8), 672-683.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development.

Applied Developmental Science, 24(2), 97-140.
<https://doi.org/10.1080/10888691.2018.1537791>

- de Bruin, A. B. H., and van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, 22, 245–252. <https://doi.org/10.1016/j.learninstruc.2012.01.003>
- Dignath, C., Büttner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review*, 3(2), 101e129. <http://dx.doi.org/10.1016/j.edurev.2008.02.003>.
- Diseth, Å. (2011). Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences*, 21(2), 191-195.
<https://doi.org/10.1016/j.lindif.2011.01.003>
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education. A review. *Studies in Higher Education*, 24(3), 331–350. <https://doi.org/10.1080/03075079912331379935>.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of Judgments of Learning (JOL) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, 33, 545–565.
<https://doi.org/10.1006/jmla.1994.1026>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest*, 5(3), 69-106. <https://doi.org/10.1111%2Fj.1529-1006.2004.00018.x>
- Duque Micán, A., & Cuesta Medina, L. (2017). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assessment & evaluation in higher education*, 42(3), 398-414.
<https://doi.org/10.1080/02602938.2015.1118433>
- Elder, A. D. (2010). Children’s self-assessment of their school work in elementary school. *Education 3-13*, 38(1), 5–11.
<https://doi.org/10.1080/03004270802602044>
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic medicine*, 80(10), S46-S54.
<https://doi.org/10.1097%2F00001888-200510001-00015>
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of educational research*, 59(4), 395-430.
<https://doi.org/10.3102%2F00346543059004395>

- Gao, M. (2009). Students' voices in school-based assessment of Hong Kong: A case study. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 107–130). Charlotte, NC: Information Age Publishing.
- García, C. (1989). *Introducción a la formación del profesorado. Teoría y métodos*. Universidad de Sevilla, Servicio de Publicaciones.
- Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: a systematic review. *Review of Education*, 5(1), 3-57. <https://doi.org/10.1002/rev3.3077>
- Gipps, C. V., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence?: Teachers and national assessment of seven year olds*. Open University Press.
- Glaser, C., Kessler, C., Palm, D., & Brunstein, J. C. (2010). Improving fourth graders' self-regulated writing skills: Specialized and shared effects of process-oriented and outcome-related self-regulation procedures on students' task performance, strategy use, and self-evaluation. *German Journal of Educational Psychology*, 24(3-4), 177–190. <https://psycnet.apa.org/doi/10.1024/1010-0652/a000015>
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of educational research*, 77(3), 334-372. <https://doi.org/10.3102/003465430303953>
- Griffiths, M., & Davies, C. (1993). Learning to learn: Action research from an equal opportunities perspective in a junior school. *British Educational Research Journal*, 19(1), 43–58. <https://doi.org/10.1080/0141192930190104>
- Harris, L. R., & Brown, G. T. (2018). *Using self-assessment to improve student learning*. Routledge.
- Harris, L. R., & Brown, G. T. L. (2010, May). "My teacher's judgment matters more than mine": Comparing teacher and student perspectives on self-assessment practices in the classroom. Paper presented to the SIG-Classroom Assessment at the American Educational Research Association Annual Conference, Denver, CO.
- Harris, L. R., Harnett, J. A., & Brown, G. T. L. (2009). 'Drawing' out student conceptions: Using pupils' pictures to examine their conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 321–330). Charlotte, NC: Information Age Publishing.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Hatzipanagos, S., & Rochon, R. (2014). *Approaches to assessment that enhance learning in higher education*. Routledge.
- Herrera-Torres, L., & Lorenzo-Quiles, O. (2009). Estrategias de aprendizaje en estudiantes universitarios. Un aporte a la construcción del Espacio Europeo de Educación Superior. *Educación y Educadores*, 12(3). <http://hdl.handle.net/10818/14944>
- Hewitt, M. P. (2005). Self-evaluation accuracy among high school and middle school instrumentalists. *Journal of Research in Music Education*, 53(2), 148. <https://doi.org/10.1177%2F002242940505300205>
- Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2018). Moving beyond P values: Everyday data analysis with estimation plots. *Nat Methods* 16, 565–566. <https://doi.org/10.1038/s41592-019-0470-3>
- House, E. R. (1993). *Professional evaluation: Social impact and political consequences*. CA: Sage Publications.
- Huang, Y., and Gui, M. (2015). Articulating teachers' expectations afore: Impact of rubrics on Chinese EFL learners' self-assessment and speaking ability. *Journal of Education and Training Studies*, 3(3), 126–132. <https://doi.org/10.11114/jets.v3i3.753>
- Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing*, 11(1), 42–65. <https://doi.org/10.1016/j.asw.2005.12.002>
- Instituto Nacional de Estadística (2020). *Mujeres graduadas en educación superior*. Disponible en https://www.ine.es/ss/Satellite?L=es_ES&c=INESeccion_C&cid=1259925481157&p=1254735110672&pagename=ProductosYServicios%2FPYSLayout¶m1=PYSDetalle¶m3=1259924822888
- Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: A study of peer- and self-assessment in a girls-only class. *Educational Studies*, 37(4), 391–403. <https://doi.org/10.1080/03055698.2010.508605>
- Jones, J. C., Trap, J., & Cooper, J. O. (1977). Technical report: Students' self-recording of manuscript letter strokes. *Journal of Applied Behavior Analysis*, 10(3), 509–514. <https://doi.org/10.1901/jaba.1977.10-509>

- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, 39(7), 840-852. <http://dx.doi.org/10.1080/02602938.2013.875117>.
- Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., & Eisenberg, S. N. (2004). Schoolage children's self-assessment of oral narrative production. *Communication Disorders Quarterly*, 26(1), 37–48. <https://doi.org/10.1177%2F15257401040260010401>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport: Praeger.
- Kasanen, K., & Rätty, H. (2002). “You be sure now to be honest in your assessment”: Teaching and learning self-assessment. *Social Psychology of Education*, 5(4), 313–328. <https://doi.org/10.1023/A:1020993427849>
- Kasanen, K., Rätty, H., & Eklund, A.-L. (2009). Elementary school pupils' evaluations of the malleability of their academic abilities. *Educational Research*, 51(1), 27–38. <https://doi.org/10.1080/00131880802704699>
- Kolić-Vehovec, S., Pahljina-Reinić, R., & Rončević Zubković, B. (2021). Effects of collaboration and informing students about overconfidence on metacognitive judgment in conceptual learning. *Metacognition and Learning*, 1-30. <https://doi.org/10.1007/s11409-021-09275-7>
- Kostons, D. (2010). *On the role of self-assessment and task-selection skills in self-regulated learning*. Open University.
- Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: a cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121–132. <https://doi.org/10.1016/j.learninstruc.2011.08.004>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Lan, W. Y. (1998). Teaching self-monitoring skills in statistics. In D. H. Schunk & B. J. Zimmerman (Eds.), *Selfregulated learning: from teaching to self-reflective practice* (pp. 86–105). New York: Guilford Press.
- Lasonen, J. (1995). A case study of student selfassessment in upper secondary education. In J. Lasonen & M.-L. Stenstrom (Eds.), *Contemporary issues of occupational education in Finland* (pp. 199–215). Jyväskylä, Finland: University of Jyväskylä, Institute for Educational Research.

- Laveault, D., & Miles, C. (2002, April). *The study of individual differences in the utility and validity of rubrics in the learning of writing ability*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- LaVoie, J. C., & Hodapp, A. F. (1987). Children's subjective ratings of their performance on a standardized achievement test. *Journal of School Psychology, 25*(1), 73–80. [https://doi.org/10.1016/0022-4405\(87\)90062-8](https://doi.org/10.1016/0022-4405(87)90062-8)
- Lázaro, M. Á. (1990). Problemas y polémicas en torno al diagnóstico pedagógico. *Bordón. Revista de pedagogía, 42*(1), 7-16.
- Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assessment & Evaluation in Higher Education, 37*(2), 137–147. <https://doi.org/10.1080/02602938.2010.515013>.
- Lee, D., & Gavine, D. (2003). Goal-setting and selfassessment in Year 7 students. *Educational Research, 45*(1), 49–59. <https://doi.org/10.1080/0013188032000086118>
- Ley Orgánica 1/1990, de 3 de octubre (BOE del 4 de octubre de 1990), de ordenación general del sistema educativo (LOGSE).
- Ley Orgánica 2/2006, de 3 de mayo (BOE del 4 de mayo), de educación (LOE).
- Ley Orgánica 9/1995, de 20 de noviembre (BOE del 21 de noviembre de 1995), de la participación, evaluación y gobierno de los centros educativos (LOPEG).
- Lopez, R., & Kossack, S. (2007). Effects of recurring use of self-assessment in university courses. *International Journal of Learning, 14*(4), 203–216. <https://doi.org/10.18848/1447-9494/CGP/v14i04/45277>
- Luyten, H., & Dolkar, D. (2010). School-based assessments in high-stakes examinations in Bhutan: A question of trust? Exploring inconsistencies between external exam scores, school-based assessments, detailed teacher ratings, and student self-ratings. *Educational Research and Evaluation, 16*(5), 421–435. <https://doi.org/10.1080/13803611.2010.530437>
- McDevitt, T. M., Sheehan, E. P., Sinco, S. R., Cochran, L. S., Lauer, D., & Starr, N. L. (2008). These are my goals: Academic self-regulation in reading by middle-school students. *Reading Improvement, 45*(3), 115–138.
- McDonald, B. (2002). Self assessment skills used by high school students without formal training. *School Psychology International, 23*(4), 416–424. <https://doi.org/10.1177%2F0143034302234004>

- McDonald, B. (2009). Exploring academic achievement in males trained in self-assessment skills. *Education 3-13*, 37(2), 145–157. <https://doi.org/10.1080/03004270802069244>
- Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica*, 113, 123–132. [https://doi.org/10.1016/S0001-6918\(03\)00023-4](https://doi.org/10.1016/S0001-6918(03)00023-4)
- Moss, C. M., & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders*. ASCD.
- Munns, G., & Woodward, H. (2006). Student engagement and student self-assessment: The REAL framework. *Assessment in Education: Principles, Policy and Practice*, 13(2), 193–213. <https://doi.org/10.1080/09695940600703969>
- Myyry, L., Karaharju-Suvanto, T., Virtala, A. M. K., Raekallio, M., Salminen, O., Vesalainen, M., & Nevgi, A. (2021). How self-efficacy beliefs are related to assessment practices: a study of experienced university teachers. *Assessment & Evaluation in Higher Education*, 1-14. <https://doi.org/10.1080/02602938.2021.1887812>
- Navarro, D.J., Foxcroft, D.R., & Faulkenberry, T.J. (2019). *Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners*. (Version 0.65). <https://learnstatswithjasp.com>
- Ndoye, A. (2017). Peer / Self Assessment and Student Learning. *International Journal of Teaching and Learning in Higher Education*, 29(2), 255-269.
- Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the “delayed-JOL effect” *Psychological Science*, 2(4), 267–270. <https://doi.org/10.1111/j.1467-9280.1991.tb00147.x>
- Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. *Review of educational research*, 53(1), 117-128. <https://doi.org/10.3102%2F00346543053001117>.
- Newton, J. M., & Nation, I. S. P. (2020). *Teaching ESL/EFL listening and speaking*. Routledge.
- Ng, J. R., & Earl, J. K. (2008). Accuracy in selfassessment: The role of ability, feedback, selfefficacy and goal orientation. *Australian Journal of Career Development*, 17(3), 39–50. <https://doi.org/10.1177%2F103841620801700307>
- Nicol, D., & McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning, a model and seven principles of good feedback practice. *Studies in*

Higher Education, 31(2), 199–218.
<https://doi.org/10.1080/03075070600572090>.

- Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research, and Evaluation*, 10(1), 17.
<https://doi.org/10.7275/a166-vm41>
- Olina, Z., & Sullivan, H. J. (2002, April). *Effects of teacher and self-assessment on student performance*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Panadero, E. (2011). Ayudas Instruccionales a la autoevaluación y la autorregulación: evaluación de la eficacia de los guiones de autoevaluación frente a la de las rúbricas. Tesis Doctoral. Departamento de Educación y Psicología del Desarrollo. Universidad Autónoma de Madrid, España.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8(422).
<http://dx.doi.org/10.3389/fpsyg.2017.00422>
- Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, 11(2), 551–576. <https://doi.org/10.14204/ejrep.30.12200>.
- Panadero, E., & Alonso-Tapia, J. (2017). Autoevaluación: connotaciones teóricas y prácticas. Cuándo ocurre, cómo se adquiere y qué hacer para potenciarla en nuestro alumnado. *Electronic Journal of Research in Education Psychology*, 11(30), 551-576. <http://dx.doi.org/10.14204/ejrep.30.12200>
- Panadero, E., & Brown, G. T. (2017). Teachers' reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156. <https://doi.org/10.1007/s10212-015-0282-5>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133e148.
<http://dx.doi.org/10.1080/0969594X.2013.877872>.
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on selfregulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806–813.
<https://doi.org/10.1016/j.lindif.2012.04.007>.
- Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers.

Studies in Educational Evaluation, 39(3), 125–132.
<https://doi.org/10.1016/j.stueduc.2013.04.001>.

- Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803-830. <https://doi.org/10.1007/s10648-015-9350-2>.
- Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, 21(4), 365-383.
<https://doi.org/10.1080/0969594X.2014.919247>
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74-98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies in Educational Evaluation*, 39(4), 195–203.
<https://doi.org/10.1016/j.stueduc.2013.10.005>.
- Papanthymou, A., & Darra, M. (2019). The Contribution of Learner Self-Assessment for Improvement of Learning and Teaching Process: A Review. *Journal of Education and Learning*, 8(1), 48-64. <https://doi.org/10.5539/jel.v8n1p48>
- Pellegrino, J. W. (2020). Important Considerations for Assessment to Function in the Service of Education. *Educational Measurement: Issues and Practice*, 39(3), 81-85. <https://doi.org/10.1111/emip.12372>
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238–250.
<https://doi.org/10.1016/j.learninstruc.2007.05.001>
- Powel, W. D., & Gray, R. (1995). Improving performance predictions by collaboration with peers and rewarding accuracy. *Child Study Journal*, 25(2), 141–154.
- Quílez-Robres, A.; Moyano, N.; Cortés-Pascual, A. (2021). Motivational, Emotional, and Social Factors Explain Academic Achievement in Children Aged 6–12 Years: A Meta-Analysis. *Education Sciences*, 11, 513.
<https://doi.org/10.3390/educsci11090513>
- Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J., & Van Gog, T. (2019). Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition and Learning*, 14(1), 21-42.
<https://doi.org/10.1007/s11409-019-09189-5>

- Raider-Roth, M. B. (2005). Trusting what you know: Negotiating the relational context of classroom life. *Teachers College Record*, 107(4), 587–628.
<https://psycnet.apa.org/record/2005-03199-003>
- Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, 20(1), 18–41.
<https://doi.org/10.4219/jaa-2008-869>.
- Ratminingsih, N. M., Marhaeni, A. A. I. N., & Vigayanti, L. P. D. (2018). Self-Assessment: The Effect on Students' Independence and Writing Competence. *International Journal of Instruction*, 11(3), 277-290.
<https://doi.org/10.12973/iji.2018.11320a>
- Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. Vintage.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and metaanalysis. *Psychological Bulletin*, 138(2), 353e387.
<http://dx.doi.org/10.1037/a0026838>.
- Rodríguez Diéguez, J. L. (2001). *La jerga de la reforma educativa*. Ariel.
- Roger Sell, G. (1989). An organizational perspective for the effective practice of assessment. *New Directions for Higher Education*, 1989(67), 21-41.
<https://doi.org/10.1002/he.36919896704>
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998). Skills training versus action research in-service: Impact on student attitudes to selfevaluation. *Teaching and Teacher Education*, 14(5), 463–477. [https://doi.org/10.1016/S0742-051X\(97\)00054-1](https://doi.org/10.1016/S0742-051X(97)00054-1)
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education: Principles, Policy & Practice*, 9(1), 81–95. <https://doi.org/10.1080/09695940220119201>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48(2), 1-36.
<http://dx.doi.org/10.18637/jss.v048.i02>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <http://dx.doi.org/10.1007/bf00117714>.

- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.
https://doi.org/10.1207/s15326977ea1101_1
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and psychological measurement, 66*(4), 701-716.
<https://doi.org/10.1177/0013164405282471>
- Schunk, D. H. (1996). Goal and self-evaluative influences during children’s cognitive skill learning. *American Educational Research Journal, 33*(2), 359–382.
<https://doi.org/10.3102/00028312033002359>
- Schuster, C., Stebner, F., Leutner, D., & Wirth, J. (2020). Transfer of metacognitive skills in self-regulated learning: an experimental training study. *Metacognition and Learning, 15*(3), 455-477. <https://doi.org/10.1007/s11409-020-09237-5>
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation practice, 17*(2), 151-161.
- Shapiro, E. S. (2014). *Behavioral assessment in school psychology*. Routledge.
- Stipek, D., & Mac Iver, D. (1989). Developmental changes in children’s assessment of intellectual competence. *Child Development, 60*, 521–538.
<https://doi.org/10.2307/1130719>
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self-and peer-assessment system. *Computers & Education, 45*(2), 187-202. <https://doi.org/10.1016/j.compedu.2004.07.002>
- Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers’ self and peer assessments. *Journal of Adolescence, 33*(1), 135–145.
<https://doi.org/10.1016/j.adolescence.2009.04.004>
- Tan, K. H. K. (2001). *Conceptions of student self-assessment. Unpublished PhD Thesis Confirmation Paper*. University of Technology, Sydney.
- Tan, K. H. K. (2012a). *Student self-assessment. Assessment, learning and empowerment*. Research Publishing.
- Tan, K. H. K. (2012b). Student self-assessment in terms of assessment. *Student self-assessment. Assessment, learning and empowerment* (pp. 15–28). Singapore: Research Publishing.

- Taras, M. (2010). Student self-assessment: processes and consequences. *Teaching in Higher Education*, 15(2), 199–209.
<https://doi.org/10.1080/13562511003620027>.
- Thiede, K.W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: an analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 25(4), 1024–1037.
<https://doi.org/10.1037/0278-7393.25.4.1024>.
- Tindal, G. A., & Marston, D. B. (1990). *Classroom-based assessment: Evaluating instructional outcomes*. Merrill Publishing Co.
- Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (Vol 1, pp. 55e87). Springer Netherlands.
- Torrance, H. & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment in the classroom*. (Buckingham, Open University Press).
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- van Helvoort, A. A. J. (2012). How adult students in information studies use a scoring rubric for the development of their information literacy skills. *The Journal of Academic Librarianship*. 38, 165–171.
<https://doi.org/10.1016/j.acalib.2012.03.016>
- Vancouver, J. B., & Day, D. V. (2005). Industrial and organisation research on self-regulation: from constructs to applications. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale*, 54(2), 155–185.
<https://doi.org/10.1111/j.1464-0597.2005.00202.x>
- Vasileiadou, D., & Karadimitriou, K. (2021). Examining the impact of self-assessment with the use of rubrics on primary school students' performance. *International Journal of Educational Research Open*, 2, 100031.
<https://doi.org/10.1016/j.ijedro.2021.100031>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48. <https://doi.org/10.18637/jss.v036.i03>
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, 34(2), 239–255.
<https://www.jstor.org/stable/canajeducrevucan.34.2.239>

- Wang, W. (2017). Using rubrics in student self-assessment: student perceptions in the English as a foreign language writing context. *Assessment & Evaluation in Higher Education*, 42(8), 1280-1292. <https://doi.org/10.1080/02602938.2016.1261993>
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: current state of the art. *Advances in Health Sciences Education*, 7(1), 63–81. <https://doi.org/10.1023/A:1014585522084>.
- Warren, E. & Nisbet, S. (1999) The relationship between the purported use of assessment techniques and beliefs about the uses of assessment, in: J. M. Truran & K. M. Truran (Eds) *22nd Annual Conference of the Mathematics Education and Research Group of Australasia* (Adelaide, SA, MERGA), 515-521.
- Watt, H. M. (2000). Measuring attitudinal change in mathematics and English over the 1st year of junior high school: A multidimensional analysis. *Journal of Experimental Education*, 68(4), 331–361. <https://doi.org/10.1080/00220970009600642>
- Wells, L., & Sweeney, P. D. (1986). A test of three models of bias in self-assessment. *Social Psychology Quarterly*, 49(1), 1–10. <https://doi.org/10.2307/2786852>
- Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, 37(1), 5–20.
- Yan, Z., Brown, G. T., Lee, J. C. K., & Qiu, X. L. (2020). Student self-assessment: Why do they do it?. *Educational Psychology*, 40(4), 509-532.
- Ysseldyke, J.E. y Matson, D. (1988). Issues in the psuchological evaluation of children. En V. Van Hasselt, P.S. Satrain, y M. Hersen (Eds.) *Handbook of developmental and physical disabilities*. Nueva York: Pergamon.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, 25(1), 82-91. <https://doi.org/10.1006/ceps.1999.1016>
- Zimmerman, B. J., & Moylan, A. R. (2009). *Self-regulation: Where metacognition and motivation intersect*. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (p. 299–315). Routledge/Taylor & Francis Group
- Zimmerman, B. J., and Schunk, D. H. (2011). “Self-regulated learning and performance: an introduction and overview,” in *Handbook of Self-Regulation of Learning and Performance*, eds B. J. Zimmerman and D. H. Schunk (New York, NY: Routledge), 1–14.

Capítulo 4: ARTICULOS CIENTÍFICOS

Por orden de presentación:

León, S. P., Pantoja-Vallejo, A., & Nelson, J. B. (2021). Variability In The Accuracy Of Self-Assessments Among Low, Moderate, And High Performing Students In University Education. *Practical Assessment, Research, and Evaluation*, 26(1), 16. <https://scholarworks.umass.edu/pare/vol26/iss1/16>

León, S. P., Pantoja-Vallejo, A., & Ferrero, M. (2021, artículo enviado). Influence of class position on academic performance and Self-Assessment: an exploratory study from a Frequentist and a Bayesian perspective. *Studies in Higher Education*.

León, S. P., & García-Martínez, I. (2021, in press). Feedback and evaluative experience as decisive factors in student self-regulation. *Publicaciones*.

León, S. P., Augusto-Landa, J. M., & García-Martínez, I. (2021). Moderating Factors in University Students' Self-Evaluation for Sustainability. *Sustainability*, 13(8), 4199. <https://doi.org/10.3390/su13084199>

León, S. P., Pantoja-Vallejo, A., & Nelson, J. B. (2021). Variability In The Accuracy Of Self-Assessments Among Low, Moderate, And High Performing Students In University Education. *Practical Assessment, Research, and Evaluation*, 26(1), 16. <https://doi.org/10.7275/6q91-az58>

ABSTRACT

The present work empirically examines the validity of Student Self-Assessment (SSA) as an educational assessment in higher education. We briefly review the principle methodological factors that could affect SSA validity, as well as the main findings identified in the literature. One empirical study is presented that compares student-self evaluations on a test with the evaluation made by the course instructor while controlling for students' experience with SSA, criteria, rubric, and scales used by the student and teacher, and that the teacher was blind. Results show a strong correlation overall between the SSA and the instructor's evaluation and show that lower-performing students tend to over-estimate their performance while higher-performing students under-estimate their performance. The results support that SSA is valid for the average student, but less so for those that deviate above and below average in the absence of measurements of potentially mediating variables. The need to consider metacognitive factors in SSA is proposed.

León, S. P., Pantoja-Vallejo, A., & Ferrero, M. (2021, artículo enviado). Influence of class position on academic performance and Self-Assessment: an exploratory study from a Frequentist and a Bayesian perspective. *Studies in Higher Education*.

ABSTRACT

Traditionally, students who are placed in the front row have been associated with better performance. In addition, it has been observed that students who perform better academically are more accurate in assessing their achievements. The objective of this study was to analyse the relationship between the seat position students choose in class, academic performance and their ability to self-evaluate. Academic performance, self-assessment scores, and confidence judgment on that assessment were analysed in a sample of 130 college students based on their seat position. In order to make a complete interpretation of the results, they were analysed from a frequentist and Bayesian approach. Results showed that seat position did not influence academic performance. On the contrary, this variable did influence students' self-assessment ability and confidence judgments on self-assessment. Students in the first rows showed overestimation and those in the last rows showed underestimation in the self-evaluation. The results are discussed according to the hypothesis of personality characteristics of students and the potential relationship between poor metacognitive skills and self-evaluation biases

León, S. P., & García-Martínez, I. (2021). Feedback and evaluative experience as decisive factors in student self-regulation. *Publicaciones*, *51*(1), 303–316. <https://doi.org/10.30827/publicaciones.v51i1.20738>

ABSTRACT

Assessment plays a key role in learning processes in educational environments. Thus, active evaluations, such as self-evaluation, are being used increasingly frequently as a supplement to traditional evaluations due to their proven formative value. The present study analyzes the impact of the evaluations made by the students on the experience acquired by these students when they are subjected to evaluation situations and when they receive feedback from the teacher on these evaluations. To this end, two natural class groups carried out evaluations of the work done by other groups and of their own work. The difference between the groups was that one group received feedback from the teacher after the assessment and the other group did not. The results were analyzed to determine the groups' accuracy in evaluating their peers and themselves with respect to the teacher's evaluation. The data showed that both groups improved the accuracy of the evaluations throughout the sessions, with a significantly greater improvement for the group that received feedback. These results highlight the importance of evaluation in the learning process and the need to provide students with feedback on their evaluation.

León, S. P., Augusto-Landa, J. M., & García-Martínez, I. (2021). Moderating Factors in University Students' Self-Evaluation for Sustainability. *Sustainability*, 13(8), 4199. <https://doi.org/10.3390/su13084199>

ABSTRACT

Background: Self-evaluation is a multidimensional construct that has raised increasing interest within educational research at different educational stages. Different studies have pointed out the important role that Student Self-Assessment plays in improving student learning and ensuring the sustainability in instructional and evaluation processes. **Method:** The aim of this study with 630 university students is to analyze how engagement profiles and study strategies (measured by questionnaire) can predict the accuracy of students' self-assessment of their achievements. For this purpose, the UWE-9 questionnaire was used to evaluate engagement, the Study Techniques Questionnaire scale to measure study strategies and a content-based test to evaluate performance, along with a self-assessment test in which the student had to estimate the level of achievement obtained in the content-based test, once the test had been completed. **Results:** The results show that both the academic engagement and the study strategies undertaken by students can be important factors that may influence different aspects of learning in the educational context. Students with higher performance and more engagement tend to show greater skills with student self-assessment (SSA) and students with better study habits tend to have better scores, greater confidence in the SSAs delivered and better skills for self-assessment. **Conclusions:** Findings suggests that providing opportunities for students to have a greater involvement in the construction of their learning and in its evaluation raises positive attitudes, which results in increased performance in order to achieve greater sustainability in the learning process assessment.