

A grade como instrumento de análise curricular na elaboração de provas padronizadas

The reticle as an instrument of curriculum analysis in the standardized tests development

La retícula como instrumento de análisis curricular en la elaboración de pruebas estandarizadas

Jorge-Gustavo Gutiérrez-Benítez*
Hiram Almejo Díaz*

Resumo

A construção de testes padronizados envolve processos metodológicos rigorosos que por sua vez orientam sua validade e confiabilidade. Nesse sentido, as primeiras etapas do desenvolvimento do teste concentram-se na determinação do universo de conhecimento que o teste pretende avaliar, processo essencial para demonstrar a validade de conteúdo e de avaliação. ao mesmo tempo justificam a representatividade do conteúdo temático. Este artigo apresenta o desenho e desenvolvimento de um instrumento conhecido como grade, para uma das disciplinas de maior relevância curricular do núcleo comum de línguas da Faculdade de Letras da Universidade Autônoma da Baixa Califórnia. Instrumento utilizado para análise curricular na construção de testes padronizados referidos ao critério, seguindo a metodologia formulada pelo Instituto de Pesquisa e Desenvolvimento Educacional da Universidade Autônoma da Baixa Califórnia, que entre outros métodos estabelece a definição de comitês de especialistas para a construção do instrumento e critérios definidos para avaliar a importância dos conteúdos temáticos envolvidos na prova. Entre os resultados destacados está a determinação do índice de relevância curricular de todos os temas que compõem o currículo da disciplina, critério utilizado para definir a representatividade de cada conteúdo na prova, bem como especificar o grau de contribuição de cada tema para o aproveitamento. de competência em cada unidade. Foram identificados conteúdos temáticos cuja relevância curricular é essencial para a concretização de conhecimentos noutras unidades de aprendizagem, bem como para conhecimentos futuros de outras disciplinas. Da mesma forma, foram identificados conteúdos de baixíssima relevância curricular,

How to cite:

Gutiérrez-Benítez, Almejo (2024) A grade como instrumento de análise curricular na elaboração de provas padronizadas. *Revista Iberoamericana De educación*, 8 (3).

Received: February, 2024
Approved: May, 2024

DOI:
<https://doi.org/10.31876/ie.v8i12.272>

<http://www.revista-iberoamericana.org/index.php/es>

* Doctor en Innovación en Tecnología Educativa
Universidad Autónoma de Baja California
Facultad de Lenguas
gutierrez.jorge@uabc.edu.mx
<https://orcid.org/0000-0003-3392-6398>

* Doctor en Administración Educativa
Universidad Estatal de Estudios Pedagógicos
dralmejo@hotmail.com,
<https://orcid.org/0000-0002-5014-2571>

denotando uma possível deficiência na construção do mapa curricular da disciplina, situação que sugere sua correção e atualização.

Palavras-chave: Padronização, Avaliação, Exames, Currículo.

Abstract

The construction of standardized tests involves rigorous methodological processes that in turn provide guidelines for the test validity and reliability. According to this, the first test development stages focus on determining the universe of knowledge that the test intends to evaluate, an essential process to demonstrate the content validity and in turn justify the representativeness of thematic content. This article presents the design and elaboration of an instrument known as the reticule, for one of the most curricularly relevant subjects in the common core of languages in the School of Languages at the Autonomous University of Baja California. Instrument used for curricular analysis in the construction of standardized referring criteria tests, following the methodology formulated by the Institute of Educational Research and Development of the Autonomous University of Baja California, which among other methods establishes the definition of expert committees for the instrument construction and defined criteria to evaluate the thematic contents importance implied in the test. Among the outstanding results is the determination of the curricular relevance index of all the topics that comprise the subject curriculum, a criterion used to define the representativeness of each content in the test, as well as to specify the degree of contribution of each topic to the achievement of the competency in each unit. Thematic contents were identified whose curricular relevance is indispensable for the knowledge acquisition in other learning units, as well as future knowledge for other subjects. Likewise, contents with very low curricular relevance were identified, indicating a possible deficiency in the subject curricular map construction, thus suggesting a necessary correction and updating.

Keywords: Standardization, Assessment, Exams, Curriculum.

Resumen

La construcción de pruebas estandarizadas conlleva procesos metodológicos rigurosos que a su vez dan pauta de la validez y confiabilidad de la misma, en ese tenor las primeras etapas del desarrollo de la prueba se enfocan en la determinación del universo

de conocimientos que la prueba pretende evaluar, proceso esencial para evidenciar la validez de contenido y a su vez justificar la representatividad de contenidos temáticos. Este artículo presenta el diseño y la elaboración de un instrumento conocido como la retícula, para una de las asignaturas de mayor relevancia curricular en el tronco común de lenguas en la Facultad de Idiomas de la Universidad Autónoma de Baja California. Instrumento empleado para el análisis curricular en la construcción de pruebas estandarizadas referidas a criterio, siguiendo la metodología formulada por el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California, que entre otros métodos establece la definición de comités de expertos para la construcción del instrumento y criterios definidos para evaluar la importancia de los contenidos temáticos implicados en la prueba. Dentro de los resultados sobresalientes se encuentra la determinación del índice de relevancia curricular de todos los temas que comprenden el currículo de la asignatura, criterio utilizado para definir la representatividad de cada contenido en la prueba, así como especificar el grado de contribución de cada tema al logro de la competencia en cada unidad. Se identificaron contenidos temáticos cuya relevancia curricular es indispensable para la concreción de conocimientos en otras unidades de aprendizaje, así como conocimientos futuros de otras asignaturas. Así mismo se identificaron contenidos que poseen muy baja relevancia curricular denotando una posible deficiencia en la construcción del mapa curricular de la asignatura, situación que sugiere una corrección y actualización del mismo.

Palabras clave: Estandarización, Evaluación, Exámenes, Currículo.

INTRODUÇÃO

O objetivo deste trabalho de investigação é apresentar o procedimento metodológico para a elaboração da grelha de um teste departamental padronizado para a disciplina de Morfologia na Língua Segunda, uma das duas disciplinas de maior relevância curricular do tronco comum de línguas da Faculdade de Letras da UABC. Detalha aspectos como a definição de critérios de avaliação específicos para determinar os índices de relevância curricular de cada uma das disciplinas que compõem todo o universo de conhecimentos abrangidos pela disciplina em questão, tais como a relevância disciplinar, a contribuição para a aquisição de competências e a contribuição para a aquisição de outros conhecimentos. São também apresentados os diferentes tratamentos

matemático-estatísticos que foram efectuados para determinar a validade de conteúdo com base no julgamento de peritos. Também se detalha o processo de justificação do conteúdo do teste e da sua representatividade, ou seja, quantos itens devem ser incluídos no teste com base na importância do conteúdo ou no seu IRC.

Posteriormente, são apresentados os resultados e produtos obtidos a partir da execução de cada um dos procedimentos propostos pela metodologia para a elaboração do teste departamental padronizado, seguidos de uma discussão e conclusão sobre as principais vantagens e contribuições proporcionadas por este instrumento da grelha na análise do currículo da disciplina em questão.

A utilização de testes padronizados para as instituições de ensino, em particular as instituições de ensino superior, é uma prática que tem sido amplamente utilizada nos últimos anos, uma vez que, dadas as exigências sociais e laborais na formação de profissionais, se opta por este tipo de avaliação para ter instrumentos válidos e fiáveis, sobre os quais se pode ter a base para a tomada de decisões com as quais se pode melhorar a qualidade da educação oferecida (Demarchi, 2020; George, 2020; Hernández et al., 2018; Muñoz Lira & Bruna Gaete, 2024).

O desenvolvimento de testes standardizados implica, na sua própria natureza, a execução de diferentes procedimentos baseados em quadros metodológicos rigorosos (Backhoff Escudero, 2018; Fernández Navas et al., 2017; Ramos Palacios et al., 2021), com os quais se pretende conferir uma qualidade educativa significativa ao instrumento de avaliação. Neste sentido, alguns autores (George, 2020; Gutiérrez Benítez & Acuña Gamboa, 2022; Jornet, 2017; Niño-Zafra, 2022) destacam entre os benefícios desse tipo de prova a possibilidade de se aproximar da realidade do domínio do conhecimento que o aluno possui pelo menos no momento da execução da prova, uma vez que os marcos metodológicos permitem apontar que a variação no desempenho alcançado na prova se deve ao tema que é avaliado, ou seja, a variação não é atribuída à qualidade técnica da prova ou ao processo de sua construção.

Na mesma linha, um dos procedimentos iniciais e importantes na elaboração dessas provas é a determinação do universo de conhecimentos que a prova pretende avaliar, e para isso é necessário implementar procedimentos específicos que permitam identificar os conteúdos temáticos que devem estar representados na prova, estabelecer prioridades entre eles, bem como indicar o impacto que esses conteúdos têm na obtenção das competências desejadas nos alunos. Isto implica, portanto, que o teste padronizado possua

validade de conteúdo (Aliaga Pacora et al., 2021; Carrillo Avalosa et al., 2020; Jiménez-Macías et al., 2021; Sánchez, 2021; Torres-Malca et al., 2022), que é entendida como a medida em que o instrumento reflecte o domínio específico do conteúdo que está a ser medido no teste.

Para alcançar o acima exposto, é comumente aplicado o método de validação por julgamento de peritos, que consiste em fazer com que diferentes especialistas na disciplina em questão emitam um juízo de valor sobre o conteúdo temático ou o universo de conhecimentos que o teste padronizado deve representar, um julgamento que, por sua vez, é tratado através de procedimentos matemático-estatísticos que permitem determinar, entre outros atributos, o índice de relevância curricular (IRC) para cada conteúdo temático, bem como o índice de concordância também conhecido como índice Kappa (Charry-Mendez et al., 2022; Garcia-Valcarcel Muñoz-Repiso et al., 2020) para os quais são utilizadas escalas de classificação, como a proposta (Landis & Koch, 1977).

Existem diferentes metodologias para a elaboração de testes normalizados, e cada uma delas propõe procedimentos específicos para a definição e validade do conteúdo de conhecimentos que o teste pretende avaliar, como é o caso da metodologia proposta pelo Instituto de Investigação e Desenvolvimento Educativo (IIDE) da Universidade Autónoma da Baixa Califórnia (UABC) para a elaboração de testes normalizados referentes a critérios, É o caso da metodologia proposta pelo Instituto de Investigação e Desenvolvimento Educativo (IIDE) da Universidade Autónoma da Baixa Califórnia (UABC) para a elaboração de testes normalizados referenciados a critérios, que estabelece a criação de um instrumento de análise curricular designado por grelha, que constitui o principal input sobre o qual se tomam as decisões que orientam a escolha dos conteúdos temáticos a incluir no teste.

MATERIAIS E MÉTODOS

A disciplina com a qual este projeto de investigação está a trabalhar é a Morfologia da Segunda Língua, que pertence à fase básica, esta disciplina é leccionada no primeiro semestre do núcleo comum de línguas, oferecido pela Faculdade de Línguas da UABC, a uma média anual de 600 alunos. Esta disciplina é uma série com a disciplina de Morfossintaxe da Segunda Língua, que é leccionada no segundo semestre.

Ambas as disciplinas são as mais relevantes curricularmente para o tronco comum, e no caso particular da Morfologia da Segunda

Língua faz parte do percurso acadêmico da Comunicação Linguística, percurso estabelecido para o desenvolvimento das competências gerais do plano de estudos da Licenciatura em Tradução, enquanto que para o plano de estudos da Licenciatura em Ensino de Línguas se articula com mais de 4 disciplinas subsequentes na fase disciplinar.

Procedimento

Como já foi referido, a grelha é um dos principais produtos gerados no âmbito da implementação da primeira fase da metodologia utilizada no desenvolvimento de exames departamentais referenciados por critérios (González-Alvarado et al., 2020; Yang, 2023) formulada pelo IIDE (Contreras Niño, 2000; Contreras Niño & Backhoff Escudero, 2004) com base no paradigma psicométrico proposto por Nitko (1994) para o desenvolvimento de testes de larga escala sob a abordagem referenciada por critérios. Estes testes são orientados pelo currículo expresso nos programas das unidades de aprendizagem (também conhecidas como cartas descritivas) das disciplinas, o que significa que os conhecimentos a avaliar, bem como a forma de os avaliar, são especificamente delimitados pelo que o currículo estabelece.

Formalmente esta metodologia estabelece seis etapas principais, no entanto, para este trabalho de investigação apenas se detalham os procedimentos envolvidos nas três primeiras etapas:

- I. Definição do domínio dos resultados pretendidos pelo currículo.
- II. Análise do currículo
- III. Desenvolvimento do plano de avaliação
- IV. Produção e validação de itens
- V. Análise primária dos resultados
- VI. Análise secundária dos resultados

Entre as primeiras ações das fases I e II, conta-se a constituição de uma comissão coordenadora de exames (CCE) e de uma comissão de conceção de exames (CDE), que por sua vez recebem formação sobre as avaliações e os procedimentos que terão de realizar. Estes comités são constituídos por diferentes professores que possuem determinados atributos desejáveis: a) conhecimentos especializados na disciplina para a qual o teste padronizado será desenvolvido, b) conhecimentos especializados no mapa curricular do núcleo comum, c) vários anos de experiência no ensino da disciplina e d) conhecimentos especializados no currículo.

A metodologia estabelece que deve ser utilizado um comité diferente para cada fase, ou seja, o mesmo professor não deve participar em

mais do que uma fase, de modo a garantir uma maior objetividade nos julgamentos efectuados. Para este trabalho de investigação, as comissões foram constituídas por professores das quatro secções estatais (Tijuana, Tecate, Ensenada e Mexicali) da Faculdade de Línguas, permitindo assim uma melhor representação da comunidade académica.

Para a Etapa I, é elaborada uma tabela de dupla entrada na qual se representam e organizam graficamente os conteúdos conceptuais e os conteúdos práticos expressos no programa da unidade de aprendizagem da disciplina de Morfologia. Posteriormente, esta tabela é apresentada a um comité para julgar e expressar, numa primeira aproximação, as relações que existem entre cada um dos conteúdos acima mencionados. Idealmente, este processo deve ser efectuado com pelo menos 3 professores que reúnam os atributos a, b e c desejáveis acima mencionados, de modo a chegar a um consenso sobre a forma como os conteúdos devem ser relacionados. Esta tabela de dupla entrada, para além de integrar todos os conteúdos do universo de conhecimento em questão, deve ainda incluir informações como a competência geral da disciplina, a competência por unidade temática, a carga horária, os possíveis percursos académicos envolvidos para a disciplina como a seriação com outra disciplina e/ou para que outras disciplinas contribui para a concretização do conhecimento.

Uma vez construída a tabela de dupla entrada, devem ser expressos verbos de ação para estabelecer, a um nível lógico, as relações previamente estabelecidas entre conteúdos conceptuais e práticos, e deve também ser utilizado um código definido para indicar os conteúdos que são essenciais, muito importantes e importantes, classificação que é atribuída uma vez concluída a segunda fase da análise curricular. O procedimento acima descrito é o que faz evoluir a grelha de dupla entrada para o que se designa por grelha. O desenho da tabela de dupla entrada e da grelha foi realizado com o software Visio, uma aplicação de diagramação e de gráficos vectoriais que faz parte do pacote de software Microsoft Office, propriedade da Microsoft.

Para a etapa II, correspondente à análise do currículo, é construída uma tabela na qual são listados todos os conteúdos conceituais e práticos da disciplina, e para cada um deles são definidos atributos de valor (ver tabela 1) que serão julgados por quatro juízes, sendo estes:

Tabela 1. *Atributos de valor a serem julgados para cada conteúdo*

Atributo	Valor
A1 Contribuição para a realização da competência	20%
A2 Dosagem (quantidade de conteúdo implícito)	10%
A3 Carga horária (atribuída ou estimada para a aprendizagem)	10%
A4 Pertinência disciplinar	20%
A5 Proporção de serviços recebidos	20%
A6 Proporção de serviços prestados	20%

Fonte: elaboração própria

Como já foi referido, a metodologia estabelece que os juízes escolhidos para esta fase não devem ser os mesmos que participaram na fase anterior. No caso do atributo Relevância Disciplinar, esse julgamento é feito por apenas um juiz, que é o especialista na disciplina, daí a denominação de juiz disciplinar; enquanto Contribuição para o alcance da competência, Dosagem e Carga horária são julgados pelos três juízes restantes. Os atributos de serviços recebidos e serviços prestados são obtidos através da contagem das relações de entrada e de saída na grelha.

Após este processo, é efectuado o cálculo estatístico do coeficiente de correlação intra-classe (ICC) através do modelo de Análise de Variância (ANOVA) para determinar a fiabilidade entre as classificações atribuídas pelos juízes. Os cálculos estatísticos acima referidos foram efectuados com o software VassarStats. A interpretação destas classificações (ver tabela 2) é dada por Landis & Koch (citados em López-Ibáñez et al., 2023) em que o índice Kappa assume valores de 0 a 1, em que 1 significa concordância perfeita.

Tabela 2. *Classificação do índice Kappa*

Valor do índice	Força
.01 - .20	Fraco - ligeiro
.21 - .40	Aceitável
.41 - .60	Moderado
.61 - .80	Considerável
.81 - 1.0	Quase perfeito

Fonte: Landis e Koch (citado em López-Ibáñez et al., 2023).

O principal produto do processo acima referido é a tabela do índice de relevância curricular (IRC) de todos os conteúdos conceptuais e práticos da disciplina. Nesta tabela, os conteúdos são ordenados

hierarquicamente com base no seu nível de importância atribuído pelo IRC.

Uma vez estabelecida a fiabilidade e a validade do julgamento, obtém-se o IRC a partir da soma dos atributos de valor julgados, estabelecendo-se um ponto de corte para classificar os conteúdos essenciais ($IRC \geq .700$), os muito importantes ($IRC \geq .400$ e $< .700$) e os importantes ($IRC < .400$ e $> .100$) e os que são objeto de avaliação quanto à sua inclusão no currículo ou considerados sem importância ($IRC \leq .100$).

Esta tabela é o principal input sobre o qual se desenvolve a fase III correspondente ao plano de avaliação, fase cujo principal produto é o documento justificativo dos conteúdos a incluir e a avaliar na prova, que toma como base o IRC de cada conteúdo para a decisão anterior, bem como fundamenta o número de itens a produzir para cada conteúdo temático.

Esta última implica que a distribuição dos itens se baseie na ponderação do IRC para o qual um item é elaborado, pelo que há mais itens na prova para os tópicos, subtópicos, etc., que têm maior relevância curricular para a concretização da competência geral do curso, bem como para os tópicos que são críticos para a realização de conhecimentos futuros.

É importante referir que a descrição da justificação de cada conteúdo deve ser muito pormenorizada, pois são suficientes, mas ao mesmo tempo devem ser apresentadas razões específicas e concretas para os atributos de valor que justificam a inclusão do conteúdo temático na prova, bem como o número de itens a testar, especificando para cada item o aspeto do conhecimento a avaliar.

RESULTADOS

Para o desenvolvimento da fase I da metodologia, o comité era composto por cinco professores, três da unidade académica de Tecate, um de Mexicali e um de Ensenada. Este comité foi responsável pela elaboração da tabela de dupla entrada, que mais tarde se tornou a grelha. Dado que, graficamente, a grelha de dupla entrada é um instrumento muito extenso, uma vez que representa todo o universo de conhecimentos envolvidos no tema, a figura 1 mostra apenas o primeiro segmento da grelha.

Figura 1. Tabla de doble entrada de morfología

<p>Competencia del curso: Identificar y aplicar los principios que se manifiestan en la formación de palabras en inglés, así como la manera en que éstas se clasifican, mediante la resolución de ejercicios escritos y orales para concienciar, y por ende mejorar, el uso de la lengua, demostrando una actitud receptiva y proactiva en el desarrollo de las sesiones.</p> <p>En particular, la materia Morfología del segundo idioma contribuye al logro de la competencia específica: <i>Manejar el idioma español y el idioma meta en el ámbito de la comunicación interpersonal, en forma oral y escrita aplicando las reglas morfológicas, fonológicas y sintácticas de una manera disciplinada y responsable.</i></p> <p>Esta materia se imparte en el primer semestre (tronco común) de las licenciaturas en Enseñanza de Lenguas y en Traducción. No recibe servicios de otro curso, pero se articula con otras asignaturas de especialización posteriores como <i>Morfosintaxis del segundo idioma, Fonología, Análisis del discurso, Lingüística general, Análisis y producción de textos en el segundo idioma.</i></p> <p>Por otra parte, forma parte de la ruta académica de <i>Comunicación Lingüística</i> que enlaza las materias de <i>Morfología del segundo idioma → Morfosintaxis del segundo idioma → Gramática avanzada en el segundo idioma → Lingüística contrastiva</i>, ruta establecida para el desarrollo de las competencias generales del Plan de Estudios de la <i>Licenciatura en Traducción.</i></p> <p>Objetivos de la evaluación: Acreditar un porcentaje de conocimiento esencial por parte del alumno.</p> <p>Uso de los resultados: Orientación al docente para mejorar su práctica.</p>	<p>Eje de contenidos conceptuales (temas y subtemas)</p>	<p>Unidad 1. Basic concepts</p> <p>Competencia: Identificar las unidades morfológicas mínimas implicadas en la formación de palabras en la lengua inglesa a través de la realización de ejercicios prácticos para crear conciencia de los mecanismos de construcción de significados; mostrando respeto, mostrando interés así como respeto hacia el profesor y sus compañeros al trabajar en el aula en equipos de trabajo o al realizar sus ejercicios de manera individual. Carga horaria: 10 horas.</p> <p>1.1 Morphology</p> <p>1.1.1 Definition of the discipline</p> <p>1.1.2 History</p> <p>1.2 Minimal meaningful units of language</p> <p>1.2.1 Free forms</p> <p>1.2.2 Signs and morphemes</p> <p>1.3 Morphology</p> <p>1.3.1 Morphemes and allomorphs</p> <p>1.3.2 Free and bound morphemes</p> <p>1.3.3 Roots, affixes</p>
	<p>Eje de contenidos prácticos (taller de morfología del segundo idioma)</p>	<p>Práctica 1</p> <p>Competencia: Identificar las unidades morfológicas mínimas implicadas en la formación de palabras en la lengua inglesa a través de la realización de ejercicios prácticos para crear conciencia de los mecanismos de construcción de significados; mostrando respeto, mostrando interés así como respeto hacia el profesor y sus compañeros al trabajar en el aula en equipos de trabajo o al realizar sus ejercicios de manera individual. Carga horaria: 6 horas</p> <p>P1 Realizar una recopilación (portafolio) de los ejercicios realizados en clase acerca de los temas analizados durante el semestre.</p>

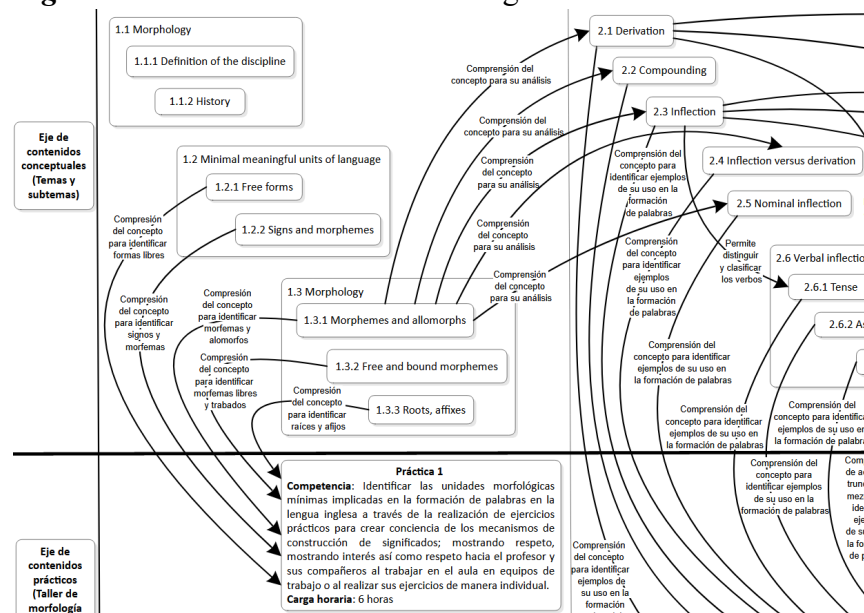
Fuente: elaboración propia

Como se puede ver en la figura 1, la tabla de doble entrada está organizada en dos ejes principales, siendo que la parte superior intermedia presenta los contenidos conceptuales e la parte inferior los contenidos prácticos. A la primera columna de la tabla se presenta la información correspondiente a la competencia general de la disciplina, al objetivo de la evaluación, a la utilización de los resultados y a la articulación de la disciplina con otras disciplinas o posibles trayectorias académicas. Una vez representados todos los contenidos curriculares previstos en el plan de estudios de la unidad de aprendizaje o en el cuadro descriptivo de la disciplina, el siguiente paso fue establecer las relaciones entre esos contenidos, de modo de identificar los contenidos que reciben servicios (conocimientos) de otros contenidos e que, por su

vez, prestam serviços a outros conteúdos. A elaboração destas relações é uma parte essencial, pois permite exprimir especificamente a um nível cognitivo qual é a ação ou a contribuição necessária de um determinado conteúdo temático em relação a outro. Ter estes níveis como referência permite ao professor ter um ponto de partida sobre o qual centrar as suas estratégias de ensino e aprendizagem, uma vez que dá uma orientação do domínio do conhecimento que o aluno deve ter em determinado conteúdo temático para que este conhecimento possa ser utilizado na concretização de outros conhecimentos, neste caso dos conteúdos temáticos com os quais existe uma relação.

Além do exposto, foi utilizado um código para indicar a importância de um conteúdo com base no IRC calculado na etapa II. A execução deste processo pelos professores que compõem a comissão transforma a tabela de dupla entrada numa primeira versão da grelha. Novamente, por se tratar de uma ferramenta gráfica de grande porte, apenas um segmento da grade e seus elementos componentes são mostrados na Figura 2.

Figura 2. Grelha do curso de Morfologia



Fonte: elaboração própria

Tal como se ilustra na figura 2, a grelha descreve graficamente como interagem todos os conteúdos temáticos expressos no currículo da disciplina, abrangendo os eixos concetual e prático. Visualmente, é possível apreciar aqueles conteúdos que são essenciais em termos do seu conhecimento, uma vez que a partir deles se estabelecem

múltiplas conexões com outros conteúdos; estas conexões são os inputs que são fornecidos para a concretização de outros conhecimentos noutros conteúdos temáticos. É importante salientar que as decisões sobre a interação entre os conteúdos resultaram do consenso de avaliação dos professores que compunham a comissão, os quais têm a experiência necessária para fazer tais juízos de valor. Dando continuidade à etapa II do desenvolvimento, a construção do quadro de conteúdos para o cálculo do índice de relevância curricular foi realizada por um novo comité, constituído por um total de quatro professores, dois da unidade de Tecate e um de Mexicali e Tijuana. O instrumento utilizado para a avaliação do currículo da disciplina é apresentado na tabela 3.

Tabela 3. *Atributos para o cálculo do IRC*

Conteúdo	A1	A2	A3	A4	A5	A6
1.1.1 Definição da disciplina	0.20	0.10	0.10	0.20	0.00	0.00
1.1.2 História	0.13	0.07	0.07	0.13	0.00	0.00
1.2.1 Formas livres	0.20	0.10	0.07	0.13	0.00	0.03
1.2.2 Sinais e morfemas	0.20	0.10	0.10	0.20	0.00	0.03

Fuente: elaboración propia

Como se pode ver na tabela 3, para cada um dos 36 conteúdos incluídos na disciplina entre conceptuais e práticos, foram avaliados os atributos acima descritos (ver tabela 1), codificando cada percentagem para o seu equivalente decimal. Após a avaliação de cada juiz, foi calculada a correlação intra-classe para determinar a concordância das avaliações emitidas pelos três juízes. O resultado deste processo é apresentado na tabela 4.

Tabela 4. *Resultados da análise do coeficiente de correlação intra-classe (ICC)*

(CCI)			
	Juez 1- Juez 2	Juez 1- Juez 3	Juez 2- Juez 3
Contribuição para a aquisição de competências	CCI = .19	CCI = .91	CCI = .33
	Tasa $r \rightarrow \text{Eta}^2 =$	Tasa $r \rightarrow \text{Eta}^2 =$	Tasa $r \rightarrow \text{Eta}^2 =$
	.59	.95	.66
	Error S $\rightarrow 1-$	Error S $\rightarrow 1-$	Error S $\rightarrow 1-$
	$\text{Eta}^2 = .40$	$\text{Eta}^2 = .05$	$\text{Eta}^2 = .34$
	ANOVA total =	ANOVA total	ANOVA total

	19.50	= 21.28	= 17.50
	Entre pares =	Entre pares =	Entre pares =
	11.50	20.28	11.50
	Intra Error =	Intra Error =	Intra Error =
	8.00	1.00	6.00
	P = .123	P = .000	P = .023
Dosagem dos conteúdos	CCI = .43	CCI = .89	CCI = .41
	Tasa r → Eta ² =	Tasa r → Eta ²	Tasa r → Eta ²
	.71	= .94	= .70
	Error S → 1-	Error S → 1-	Error S → 1-
	Eta ² = .30	Eta ² = .06	Eta ² = .30
	ANOVA total =	ANOVA total	ANOVA total
	27.50	= 26.88	= 24.88
	Entre pares =	Entre pares =	Entre pares =
	19.50	25.38	17.38
	Intra Error =	Intra Error =	Intra Error =
	8.00	1.50	7.50
	P = .003	P = .000	P = .005
Carga horária	CCI = .22	CCI = .81	CCI = .44
	Tasa r → Eta ² =	Tasa r → Eta ²	Tasa r → Eta ²
	.60	= .90	= .72
	Error S → 1-	Error S → 1-	Error S → 1-
	Eta ² = .40	Eta ² = .10	Eta ² = .3
	ANOVA total =	ANOVA total	ANOVA total
	40.00	= 30.00	= 35.28
	Entre pares =	Entre pares =	Entre pares =
	24.00	27.00	25.28
	Intra Error =	Intra Error =	Intra Error =
	16.00	3.00	10.00
	P = .101	P = .000	P = .000

Fuente: elaboración propia

Os resultados do ICC expressos na tabela acima encontram-se dispersos pelas diferentes categorias de interpretação propostas por Landis e Koch, obtendo-se algumas concordâncias ligeiras (.19), regulares (.22 e .33), moderadas (.41, .43 e .44) e quase perfeitas (.81, .89 e .91). Por este motivo, foram utilizados os juízos de valor expressos pelos juízes 1 e 3, uma vez que foram os que apresentaram

o índice de concordância Kappa quase perfeito.

Uma vez cobertas a validade e a fiabilidade dos juízos de valor, calculou-se o IRC para cada um dos 36 conteúdos, resultado da soma de cada um dos seis atributos avaliados, expressos numa escala de 0 a 1, em que quanto mais próximo de um for o IRC, maior é a importância, e o inverso se for próximo de zero. Daqui resultou a tabela IRC para a disciplina de Morfologia (ver quadro 5), que permite identificar e organizar por relevância curricular todo o universo de conhecimentos que se pretende abordar na prova normalizada.

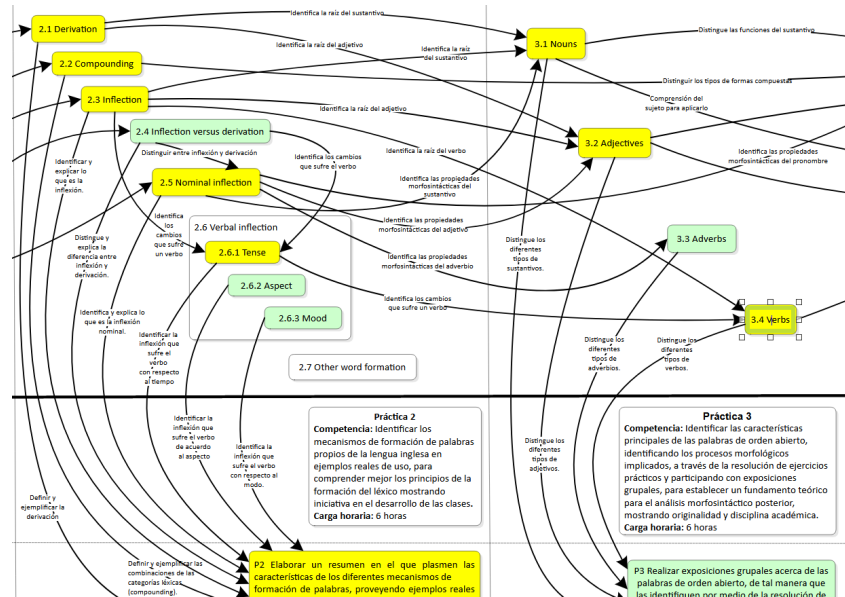
Tabela 5. *Conteúdos temáticos por índice de relevância curricular*

Conteúdo	Atributos						IRC
	A1	A2	A3	A4	A5	A6	
2.5 Flexão nominal	0.20	0.10	0.10	0.20	0.05	0.17	0.817
P2 Elaborarán un resumen en	0.20	0.10	0.10	0.20	0.20	0.00	0.800
2.3 A inflexão	0.20	0.10	0.10	0.20	0.03	0.17	0.792
1.3.1 Morfemas e alomorfos	0.20	0.10	0.08	0.20	0.00	0.20	0.783
3.1 Substantivos	0.20	0.10	0.10	0.20	0.08	0.10	0.775
3.2 Adjectivos	0.20	0.10	0.07	0.20	0.08	0.10	0.742
2.1 Derivação	0.20	0.10	0.10	0.20	0.03	0.10	0.725
2.6.1 Tempo verbal	0.20	0.10	0.10	0.20	0.05	0.07	0.717

Fonte: elaboração própria

A tabela 5 apresenta um segmento de todos os conteúdos temáticos da disciplina e o respetivo IRC. Foram estabelecidos quatro pontos de corte com base na distribuição dos valores do IRC, que permitem codificar na grelha os conteúdos que eram essenciais, os mais importantes, os importantes e os sem importância. Esta codificação e a versão final da grelha podem ser observadas na figura 3, que distingue a importância do conteúdo com uma cor específica, para que a classificação do conteúdo com base na sua importância dada pelo IRC possa ser visualmente apreciada.

Figura 3. Grelha da disciplina Morfologia codificada de acordo com o IRC



Fuente: elaboración propia

A fase II terminou com a referida atualização da grelha, pelo que se prosseguiu com a fase III, correspondente ao desenvolvimento do plano de avaliação, fase que teve como principal objetivo a elaboração do quadro justificativo (ver Quadro 6) dos conteúdos importantes a avaliar no exame normalizado departamental. Esta fase foi realizada por um novo comité, composto por três professores, um de cada unidade de Tecate, Mexicali e Tijuana.

Tabla 6. Justificación de contenidos a evaluar en el examen

Conteúdo a avaliar no exame	Importância atribuída	Razões que justificam a decisão	Observações
	Muito importante	O conteúdo 1.1.1 Definição da disciplina tem um IRC = 0,600, num máximo possível de 1,000. Embora não forneça nem receba serviço de outros conteúdos do programa, a sua importância reside no facto de ter recebido as	O caderno de encargos prevê a elaboração de dois itens, um para testar o domínio do conceito de morfologia por definição e o outro para testar a capacidade de identificar exemplos de análise morfológica.

pontuações mais elevadas em todos os critérios que compõem o Índice de Relevância Curricular (IRC): contribuição para a realização da competência da unidade, quantidade de conteúdo implícito, carga de aprendizagem estimada e relevância disciplinar.

Fuente: elaboración propia

O quadro acima exemplifica o tratamento efectuado para a justificação dos conteúdos temáticos que devem ter representatividade na prova, os critérios em que se basearam as decisões tomadas pela comissão são dados pelo IRC, bem como as valorizações obtidas em cada um dos atributos de valor que foram julgados para cada conteúdo, entre os quais se faz alusão aos serviços (conhecimentos) que presta a outros conteúdos ou aos que recebe.

Por exemplo, o conteúdo 1.3.1 Morfemas e alomorfos foi considerado um conteúdo essencial por ter um IRC muito elevado, ou seja, foi o quarto conteúdo em ordem de importância dada pelo IRC, uma vez que obteve as pontuações mais elevadas em termos de contribuição para a realização da competência da unidade (0,2), quantidade de conteúdo implícito (0,1) e relevância disciplinar (0,2). Além disso, este conteúdo temático era necessário para a concretização do conhecimento de cinco conteúdos temáticos conceptuais e um prático.

Face ao exposto, a comissão considerou a elaboração de quatro itens para representar este conteúdo temático na prova, testando o domínio do conceito de Morfema e Alomorfos, bem como a capacidade de identificar exemplos destes num texto autêntico. Estes itens centrar-se-iam nos níveis cognitivos de compreensão e aplicação, respetivamente.

CONCLUSÕES

Falar da conceção de testes padronizados implica dar orientações para procedimentos metodológicos rigorosos, com os quais se pretende limitar as possibilidades de erro associadas ao teste na avaliação do domínio de conhecimentos dos examinandos. No

entanto, uma parte importante da construção destes testes assenta na definição do conteúdo temático que o teste deve representar, sendo este um fator essencial a ter em consideração pelo examinando, uma vez que se pressupõe que o teste deve incluir os tópicos que foram leccionados nas aulas, pois, se assim não for, o desempenho no teste fica diretamente condicionado.

Na execução da proposta de metodologia para a elaboração de testes padronizados analisada neste trabalho, as duas primeiras fases incidiram, respetivamente, na definição do domínio de conhecimentos pretendido pelo teste, bem como na análise concreta do currículo da disciplina em causa.

O produto resultante da primeira etapa foi a grelha, na qual foi possível observar uma distribuição gráfica de todos os conteúdos temáticos, tanto conceptuais como práticos, que compõem o universo de conhecimentos da disciplina de morfologia; esta organização visual permitiu estabelecer mais facilmente as relações entre os conteúdos e, ao mesmo tempo, identificar os conteúdos que forneciam conhecimentos necessários para atingir os conhecimentos dos conteúdos subsequentes.

Os conectores utilizados nas relações entre um conteúdo e outro são uma vantagem substancial da grelha, uma vez que proporcionam uma imagem mais objetiva da interação entre os conteúdos temáticos e do enfoque que o professor deve aplicar na sua abordagem. Por conseguinte, é importante que a construção destas relações e verbos de ação envolva a participação de peritos e, por sua vez, um processo de validação dos juízos feitos por estes.

Para a segunda fase, o principal produto gerado foi a tabela com o IRC, que foi obtido a partir da avaliação de determinados critérios específicos definidos como parte da metodologia proposta, este processo foi crítico para a análise curricular da disciplina de morfologia, pelo que foi necessário um procedimento rigoroso para dar uma orientação da validade do IRC, razão pela qual foi efectuado o cálculo do índice de concordância Kappa para determinar a proporção de concordância em relação aos julgamentos feitos pelos três juízes especialistas. Desta forma, foi possível ter a certeza das avaliações emitidas e, com base nelas, estabelecer, por ordem de importância curricular, os conteúdos mais relevantes, que deveriam ser mais representativos no teste.

Este último aspeto reflectiu-se na terceira fase, correspondente ao plano de avaliação, uma vez que foi aqui que se definiram todos os itens a desenvolver num instrumento designado por tabela de justificação de conteúdos da prova. Como referido anteriormente, a

TCI é a principal base sobre a qual se decide qual o conteúdo temático que deve estar representado no teste, bem como a proporção desse conteúdo. O principal contributo desta metodologia para a elaboração de testes normalizados reside precisamente no que atrás foi referido, em poder estabelecer objetivamente e com fundamentos válidos o domínio do conhecimento expresso no teste, o qual está diretamente relacionado com o currículo estabelecido no programa da unidade de aprendizagem da disciplina, proporcionando assim a fiabilidade de que o teste será efetivamente constituído pelos conteúdos temáticos que foram estabelecidos desde o início na criação da disciplina em questão e qualquer decisão sobre a inclusão ou não dos itens é justificada e validada pelo respetivo IRC.

A análise curricular da disciplina de morfologia utilizando a metodologia proposta neste trabalho de investigação permitiu observar alguns aspectos em relação a determinados conteúdos temáticos. Por exemplo, os juízes identificaram dois temas com os níveis de IRC mais baixos de todo o currículo, que não tiveram impacto na aquisição de conhecimentos ou na realização de alguns subsequentes, um deles encontrava-se na unidade dois, que apresentava o maior número de conteúdos com IRC mais elevado, e o outro na unidade cinco.

Este tipo de informação é muito importante para os processos de atualização curricular, nomeadamente quando é necessário atualizar os programas das unidades de aprendizagem, uma vez que, com base nestas análises, podem ser tomadas melhores decisões na reestruturação dos domínios de resultados pretendidos pelo currículo, quer eliminando, quer acrescentando conteúdos temáticos. Por outro lado, embora este trabalho de investigação tenha incluído o desenvolvimento das primeiras etapas da elaboração de um teste normalizado, a aplicação destes mesmos procedimentos metodológicos para a elaboração de testes não normalizados tem os mesmos benefícios, principalmente o de dispor de informação válida e fiável sobre os conteúdos temáticos e a sua representatividade nos testes.

Do ponto de vista do ensino e do processo de avaliação, uma das situações mais recorrentes em que o processo de avaliação comumente se inicia é a resposta a várias incógnitas, tais como que conteúdos avaliar, quantas perguntas, itens ou reagentes devem ser feitos sobre um determinado tema, como saber qual o tema mais importante, etc. Mais uma vez, é na resposta a estas questões que importa saber quantas perguntas, itens ou reagentes devem ser feitos sobre um determinado tema, e como saber qual o tema mais

importante. Mais uma vez, é na resposta a estas questões que reside a principal contribuição deste trabalho de investigação: os métodos aqui analisados fornecem aos professores um compêndio de instrumentos com os quais podem obter uma imagem precisa do universo de conhecimentos de um determinado assunto.

Ao observar graficamente, através da grelha, as interações entre todos os conteúdos temáticos conceptuais e práticos, identificando quantitativamente o índice de relevância curricular de cada um desses conteúdos, o que fornece informação sobre a qual se pode efetivamente decidir quais as matérias que devem ser obrigatoriamente representadas na prova, bem como a proporção de itens dessa matéria, decisões que ficam registadas na tabela de justificação dos conteúdos a avaliar na prova.

Por outro lado, a existência deste tipo de análise curricular oferece também a possibilidade de fornecer um feedback mais detalhado sobre o desempenho do aluno no teste padronizado, uma vez que não só fornece informação sobre o nível geral de domínio que o aluno apresenta no momento da realização do teste, mas também sobre o nível de desempenho nas disciplinas de maior relevância curricular, o que permite orientar melhor o percurso educativo do aluno e concentrar esforços na melhoria do domínio de conhecimentos específicos.

E, na mesma linha, outra das contribuições da aplicação destes métodos de análise curricular tem a ver com a própria prática do professor, já que conhecer especificamente os temas em que a relevância curricular é maior, os temas que recebem uma maior quantidade de conhecimentos de outros temas e, por sua vez, aqueles que proporcionam conhecimentos para a realização de outros, permite-lhe concentrar os seus esforços numa melhor aplicação de estratégias de ensino e aprendizagem desses temas, com vista a que o aluno alcance um melhor domínio dos mesmos.

No entanto, o facto de a aplicação destes métodos ter identificado conteúdos temáticos que têm uma relevância curricular muito baixa ou quase nula realça a importância de estabelecer periodicamente procedimentos de controlo de rotina com os quais se possa validar adequadamente a relevância dos conteúdos temáticos nos programas das unidades de aprendizagem das disciplinas que compõem os diferentes programas de ensino, embora se saiba que a elaboração destes programas de unidades de aprendizagem implica procedimentos e métodos muito rigorosos, considerando que a aplicação destes instrumentos de análise curricular contribuiria para melhorar a sua qualidade.

Como linha de trabalho futuro, seria necessário estabelecer um esquema de formação para os professores que lecionam esta disciplina, para que se familiarizem com o currículo da disciplina, utilizando os instrumentos gerados neste trabalho de investigação, e para que possam utilizá-los para planejar melhor as estratégias didáticas a implementar na lecionação do currículo envolvido na disciplina, em particular os conteúdos temáticos com maior IRC.

Seria também importante a elaboração de um relatório de feedback padronizado que evidencie detalhadamente os níveis de desempenho que o aluno obteve na prova, para que o aluno tenha informação sobre os tópicos curriculares mais relevantes que recebeu durante o curso e assim identificar as áreas de conhecimento em que é competente e em que precisa de melhorar, bem como conhecer as possíveis implicações do desempenho alcançado na disciplina para a realização de futuros conhecimentos noutras disciplinas ou como parte de possíveis percursos académicos. Isto permitiria ao aluno saber de forma mais objetiva como a nota obtida no teste é medida em termos dos domínios de conhecimento de cada uma das disciplinas que compõem o currículo da disciplina.

Na mesma linha, seria benéfico um instrumento de feedback para o professor, no qual se pudesse apresentar informação sobre os níveis médios ou gerais de domínio alcançados pelos alunos nos tópicos de maior relevância curricular, bem como o comportamento ao nível das unidades temáticas, da mesma forma que se doseia os níveis de domínio ao nível dos itens, dado que a cada um deles está associado um conteúdo temático específico, Isto forneceria ao professor informação suficientemente abstrata para orientar melhor a sua prática pedagógica em termos da escolha de estratégias pedagógicas para tratar os temas previamente identificados como tendo um IRC mais elevado, bem como aqueles em que se observou um baixo nível de domínio.

Este tipo de análise curricular permite uma tomada de decisão mais objetiva, válida e fiável sobre o que deve ser avaliado num teste, graças a uma visão completa do universo de conhecimentos com que se está a trabalhar, fornecendo assim uma base sólida para os professores orientarem ou adaptarem a sua prática. A capacidade de classificar e hierarquizar adequadamente estes conhecimentos com base no seu índice de relevância curricular coloca o professor sempre na direção certa em termos curriculares, permitindo uma utilização mais eficiente dos seus esforços nas estratégias de ensino e aprendizagem que contribuem para a concretização dos

conhecimentos mais relevantes para o desenvolvimento das competências profissionais desejadas.

REFERENCES

- Aliaga Pacora, A. A., Juárez Hernández, L. G., y Herrera Meza, R. (2021). Diseño y validez de contenido de una rúbrica analítica socioformativa para evaluar competencias investigativas en posgrado. *Apuntes Universitarios*, 11(2), 62–82. <https://doi.org/10.17162/au.v11i2.632>
- Backhoff Escudero, E. (2018). Evaluación estandarizada de logro educativo: Contribuciones y retos. *Revista Digital Universitaria*, 19(6). <https://doi.org/10.22201/codeic.16076079e.2018.v19n6.a3>
- Carrillo Avalosa, B. A., Sánchez Mendiola, M., y Leenen, I. (2020). El concepto moderno de validez y su uso en educación médica. *Investigación en Educación Médica*, 9(33), 98–106. <https://doi.org/10.22201/facmed.20075057e.2020.33.19216>
- Charry-Mendez, S., Cabrera-Díaz, E., y Reyes-Gaspar, P. L. (2022). Análisis de concordancia inter-jueces para medir factores protectores en salud en universitarios. *Revista Ciencia y Cuidado*, 19(1), 95–106. <https://doi.org/10.22463/17949831.3055>
- Demarchi, G. (2020). La evaluación desde las pruebas estandarizadas en la educación en Latinoamérica. *Contexto Revista de Investigación en Administración, Contabilidad, Economía y Sociedad*, 8(13), 107–133.
- Fernández Navas, M., Alcaraz Salarirche, N., y Sola Fernández, M. (2017). Evaluación y Pruebas Estandarizadas: Una Reflexión sobre el Sentido, Utilidad y Efectos de estas Pruebas en el Campo Educativo. *Revista Iberoamericana de Evaluación Educativa*, 10.1(2017). <https://doi.org/10.15366/riee2017.10.1.003>
- García-Valcarcel Muñoz-Repiso, A., Hernández Martín, A., Martín del Pozo, M., y Olmos Migueláñez, S. (2020). Validación de una rúbrica para la evaluación de trabajos fin de máster. *Profesorado, Revista de Currículum y Formación del Profesorado*, 24(2), 72–96. <https://doi.org/10.30827/profesorado.v24i2.15151>
- George, C. (2020). Pruebas estandarizadas y calidad de la educación en México, sexenio 2012-2018. *Revista Universidad y Sociedad*, 12(4), 418–425.
- González-Alvarado, F., Acosta-Ballesteros, I., Artavia-Aguilar, C. V., Calderón-Ramírez, G., Chaverri-Chaves, P., Cruz-Arroyo, E., Ramírez-González, A., Rodríguez-Pineda, M., y Solano-Alpizar, J. (2020). Hallazgos principales de la investigación “13-096-19-Balance crítico de las Pruebas Nacionales de Bachillerato en

- Educación Media como instrumento de certificación del conocimiento del estudiantado”. *Revista Electrónica Educare*, 24(1), 1–24. <https://doi.org/10.15359/ree.24-1.13>
- Gutiérrez Benítez, J. G., y Acuña Gamboa, L. A. (2022). Evaluación estandarizada de los aprendizajes: una revisión sistemática de la literatura. *CPU-e, Revista de Investigación Educativa*, 34. <https://doi.org/10.25009/cpue.v0i34.2800>
- Hernández, M., Ramírez, É., y Gamboa, S. (2018). La implementación de una evaluación estandarizada en una institución de educación superior. *Innovación Educativa*, 18(76), 149–170.
- Jiménez-Macías, I. U., César Vázquez-González, G., Juárez-Hernández, L. G., y Bracamontes-Ceballos, E. (2021). Inventario de Habilidades Socioemocionales y Salud Mental para Profesores de Educación Superior: validez de contenido. *Revista Fuentes*, 2(23), 204–220. <https://doi.org/10.12795/revistafuentes.2021.12052>
- Jornet, J. (2017). Evaluación estandarizada. *Revista Iberoamericana de Evaluación Educativa*, 10(1), 5–8.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- López-Ibáñez, F. J., Cascales Martínez, A., y Martínez-Segura, M. J. (2023). Diseño universal para el aprendizaje y TIC en el área de educación física: diseño y validación de una propuesta de intervención. *PUBLICACIONES*, 53(3), 135–178. <https://doi.org/10.30827/publicaciones.v53i3.23867>
- Muñoz Lira, M. S., & Bruna Gaete, J. A. (2024). Assessment Processes and Standardized Tests. Are they compatible if we look for quality in education? *Región Científica*. <https://doi.org/10.58763/rc2024204>
- Niño-Zafra, L. S. (2022). Pruebas estandarizadas Saber y evaluaciones aula: ¿relaciones de proximidad o distanciamiento? *Área de ciencias sociales. Nodos y Nudos*, 87–98. <https://doi.org/10.17227/nyn.vol7.num52-13758>
- Ramos Palacios, L. A., Guifarro, M. I., y Casas García, L. M. (2021). Dificultades en el aprendizaje del álgebra, un estudio con pruebas estandarizadas. *Bolema: Boletim de Educação Matemática*, 35(70), 1016–1033. <https://doi.org/10.1590/1980-4415v35n70a21>
- Sánchez, R. (2021). El tema de validez de contenido en la educación y la propuesta de Hernández-Nieto. *Latin-American Journal of Physics Education*, 15(3), 1–5.
- Torres-Malca, J. R., Vera-Ponce, V. J., Zuzunaga-Montoya, F. E., Talavera, J. E., y De La Cruz-Vargas, J. A. (2022). Validez de contenido por juicio de expertos de un instrumento para medir

conocimientos, actitudes y prácticas sobre el consumo de sal en la población peruana. *Revista de la Facultad de Medicina Humana*, 22(2), 273–279. <https://doi.org/10.25176/RFMH.v22i2.4768>

Yang, Q. (2023). Should Standardization Tests be Used to Assess Student Ability? *Journal of Education, Humanities and Social Sciences*, 8, 319–326. <https://doi.org/10.54097/ehss.v8i.4268>