

Univerzita Karlova

Filozofická fakulta

Ústav románských studií

Bakalářská práce

Zdeňka Berkovcová

Druhy hodnocení ve výuce cizích jazyků v českých školách

Types of assessment in the teaching of foreign languages in Czech schools

Tipos de evaluación en la enseñanza de lenguas extranjeras en las escuelas

checas

Poděkování

Děkuji vedoucímu bakalářské práce Mgr. et Mgr. Danieli Alejandru Márquezi Guzmánovi, Ph.D. za podnětné rady a metodickou a odbornou pomoc při zpracování mé práce. Zároveň děkuji respondentům, kteří se v rámci výzkumu podělili o své zkušenosti s hodnocením ve výuce jazyků.

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Abstrakt

Tato práce se zabývá typy hodnocení ve výuce cizích jazyků na českých školách. Cílem je teoreticky představit, jaká hodnocení lze ve výuce použít, a následně zjistit, zdali je učitelé využívají, jaké názory na ně mají a jak se do jejich přístupu k hodnocení promítají vnější faktory jako např. zvyklosti ve vzdělávacím systému.

K řešení byl použit řízený rozhovor, který vycházel ze souboru tematických otázek. Respondenti byli gymnaziální učitelé.

Provedeným výzkumem bylo zjištěno, že učitelé používali jak sumativní, tak formativní hodnocení. Z formativního hodnocení pro ně bylo důležité dávat zpětnou vazbu studentům a používat pro hodnocení aktivity, které reflektují používání jazyka v běžném životě. Málo nebo vůbec používali sebehodnocení a vzájemné hodnocení studentů. Dynamické hodnocení nepoužívali. Z tlaků okolí pro ně byl nejvýraznější tlak vzdělávacího systému na používání známek.

Hlavním zjištěním práce je, že učitelé používají celou řadu hodnocení a hledají kompromisy mezi požadavky vzdělávacího systému (např. používání známek) a inovativnějšími přístupy jako je zpětná vazba a hodnocení založené na úkolech (task-based language assessment).

Klíčová slova: hodnocení ve výuce jazyků, sumativní hodnocení, formativní hodnocení, dynamické hodnocení, hodnocení založené na úkolech, sebehodnocení, vzájemné hodnocení studentů, portfolia, jazyky v českém vzdělávacím systému

Abstract

This thesis deals with the types of assessment in foreign language classes in Czech schools. The aim is to provide a theoretical introduction to the variety of assessments that can be used, and then investigate whether teachers use them, what they think about them and how their approach to assessment is influenced by aspects such as the common practices in the educational system.

To obtain the results a semi-structured interview has been used based on a series of questions about assessment. The participants of the study were grammar-school teachers.

It has been found that the teachers used both summative and formative assessments. In the latter, they put emphasis on providing their students with feedback and using activities which reflected real-life situations. They hardly ever or never used self-assessment and peer-assessment. No one used dynamic assessment. The most significant outer influence on their assessment was the pressure to use marks within the Czech educational system.

The main finding of the thesis is that Czech teachers use a range of assessments and try to find balance between the requirements of the school system (for example, the use of marks) and more innovative approaches, such as giving feedback or assessing the performance on tasks based on real life (task-based language assessment).

Key words: assessment in language learning, summative assessment, formative assessment, dynamic assessment, task-based language assessment, self-assessment, peer-assessment, portfolios, languages in the Czech educational system

Obsah

Obsah	7
Introducción	5
1. Marco teórico	5
1.1 La evaluación	5
1.1.1 El surgimiento de la evaluación de lenguas	5
1.1.2 Definiciones de la evaluación	6
1.1.3 La validez y el <i>construct</i>	7
1.1.4 La fiabilidad y el impacto de la evaluación	8
1.1.5 La ética de la evaluación	8
1.2 La evaluación a gran escala	9
1.2.1 Definición y <i>maturita</i>	9
1.2.2 Beneficios	10
1.2.3 La influencia en la enseñanza	10
1.2.4 Límites de la evaluación a gran escala	11
1.3 La evaluación en clase	12
1.3.1 Definición de la evaluación en clase	12
1.3.2 Los <i>constructs</i>	13
1.3.4 Las creencias de los profesores	13
1.3.6 Diferencias entre la evaluación a gran escala y la evaluación en clase	14
1.4 La evaluación sumativa	16
1.5 La evaluación formativa	18

1.5.1	Introducción y definición	18
1.5.2	La evaluación formativa en el aula	19
1.5.3	Beneficios.....	21
1.5.4	Problemas con la evaluación formativa	22
1.5.5	Otras formas de la evaluación formativa	23
1.6	Evaluación basada en tareas	24
1.6.1	Introducción y definición	24
1.6.2	Enfoques en TBLA.....	24
1.6.3	Beneficios de TBLA.....	25
1.6.4	Limitaciones de TBLA.....	26
1.7	La evaluación dinámica	27
1.7.1	Vygotsky y la Zona de Desarrollo Próximo	27
1.7.2	Beneficios y limitaciones de la evaluación dinámica	28
1.8	La autoevaluación	30
1.8.1	Característica de la autoevaluación	30
1.8.2	Beneficios de la autoevaluación.....	31
1.8.3	Limitaciones de la autoevaluación	31
1.9	La evaluación entre pares	33
1.9.1	Definición	33
1.9.2	Beneficios.....	33
1.9.3	Limitaciones.....	33
1.10	Portafolios.....	35

1.10.1 Definición de portafolio	35
1.10.2 Beneficios y limitaciones	35
2 Metodología	37
2.1 El estudio.....	37
2.2 Participantes	38
2.3 Instrumentos	40
2.4 Procedimiento	40
2.5 Análisis de datos.....	41
3. Resultados.....	41
3.1 Experiencia general con la evaluación	41
3.2 Evaluación a gran escala: <i>maturita</i>	42
3.3 Evaluación en clase	43
3.4 Evaluación formativa	44
3.5 Evaluación sumativa	45
3.6 Evaluación basada en tareas.....	46
3.7 Evaluación dinámica	46
3.8 Discusión.....	46
3.9 Limitaciones del estudio y vías de investigación	49
4 Conclusión	49
Anexos	52
Anexo 1	52
Referencias.....	56

Introducción

Este trabajo trata de los tipos de evaluación en la enseñanza de lenguas en las escuelas checas. La evaluación que podemos usar en el aula se divide en dos tipos básicos: la sumativa y la formativa. La formativa incluye una variedad de métodos y actividades. Por ejemplo, entre los tipos formativos contamos la evaluación dinámica o basada en tareas, la creación de portafolios, la autoevaluación o la evaluación entre pares. El objetivo de esta tesis es presentar los tipos de evaluación y luego determinar si los profesores los usan, qué piensan sobre ellos y si hay otros participantes en el proceso de aprendizaje (como el sistema educativo o los padres de los estudiantes) quienes influyen en la evaluación.

1. Marco teórico

1.1 La evaluación

1.1.1 El surgimiento de la evaluación de lenguas

Las pruebas oficiales de lenguas extranjeras se difundieron durante la primera guerra mundial y la época de posguerra cuando el gobierno y el ejército estadounidenses quisieron comprobar si los empleados potenciales eran capaces de aprender idiomas. Para hacer esto usaron *the Modern Language Aptitude Test*. En el Reino Unido también había pruebas, pero el objetivo era comprobar si los profesores tenían suficientes competencias lingüísticas para enseñar lenguas (Tsagari et al., 2016).

Tanto en la preparación para estas pruebas como en las aulas, la enseñanza y la evaluación se dirigían por el Método de Gramática-Traducción, que se enfocaba en traducciones de lengua

materna (L1) a segunda lengua (L2) y viceversa. Se ponía énfasis en traducir sin errores, no en comunicar con fluidez (Turnbull, 2020). Luego, la actitud cambió y tanto en la enseñanza como en la evaluación se destacaba la necesidad de comunicar. En 1980, Canale y Swain introdujeron la enseñanza comunicativa de la lengua (*the Communicative Approach*) que destaca el aprendizaje de la lengua para comunicar significado, es decir, combina los conocimientos de la gramática y el léxico con la pragmática, las competencias sociolingüísticas, etc. (Tzagari et al., 2016).

Después de la introducción de la enseñanza comunicativa, el Consejo de Europa desarrolló el Marco Común Europeo de Referencia (MCER) (Tzagari et al., 2016). Se divide en seis niveles de competencias, desde los más básicos (A1, A2) hasta los de usuarios independientes (B1, B2) y expertos (C1, C2). Cada nivel contiene descriptores de lo que el usuario debería saber hacer (Cummins et al., 2009). El objetivo de este marco es ofrecer un punto de referencia para la creación de currículos, evaluaciones y materiales de estudio, y unir el sistema de la enseñanza de lenguas en Europa (Tzagari et al., 2016). Tener un marco común promueve la movilidad entre países europeos porque a pesar de que haya diferencias entre los sistemas educativos, existe un marco compartido por todos al cual la gente puede referirse cuando describe sus habilidades. El uso del MCER también apoya a la alfabetización evaluativa (Purpura, 2016).

1.1.2 Definiciones de la evaluación

Hay varias definiciones de evaluación. Por ejemplo, según Black y William (1998), son todas las actividades que apoyan el progreso del estudiante (Tzagari et al., 2016). De acuerdo con Brown (2003), la evaluación es una colección de información sobre cómo el estudiante puede desempeñar una tarea (Ismail et al., 2022). Generalmente podemos decir que la evaluación lingüística nos brinda datos relacionados con el desempeño lingüístico del estudiante. Cada evaluación usa una actividad que trata de suscitar un rendimiento. Si hay

consistencia en estos rendimientos, se usa para crear datos que conducen al otorgamiento de una puntuación, una retroalimentación verbal u otro tipo de resultado. Todo esto luego se interpreta para tomar decisiones sobre el futuro de la enseñanza y el aprendizaje (Purpura, 2016). La evaluación es una parte clave de la enseñanza porque sin ella no sabemos si el estudiante ha logrado los objetivos del aprendizaje (Ismail et al., 2022).

1.1.3 La validez y el *construct*

Bachman y Palmer (2010) afirman que la evaluación lingüística debería tener un objetivo y debería basarse en un contenido determinado, como una unidad en el libro de texto o un modelo de competencia. El objetivo puede ser, por ejemplo, de diagnóstico en cuanto al progreso o dominio o proponer información sobre qué debería ser enseñado en las clases siguientes (Purpura, 2016). La problemática del contenido definido está relacionada con la validez de la evaluación. La validez significa que la prueba evalúa exactamente lo que hemos querido evaluar (Hamps-Lyons, 2009). Por eso, es necesario definir precisamente el *construct* que queremos evaluar cuando estamos desarrollando la evaluación para que tengamos un punto de referencia cuando comprobamos la validez (de Jong et al., 2016).

Hay distintas maneras de definir un *construct*. Primero, en el enfoque basado en elementos particulares (*the trait-based approach*) decidimos qué elemento queremos medir. Luego, elegimos tareas que suscitarán un desempeño con este elemento. Suponemos que, si el estudiante puede mostrar el uso de este elemento en la tarea concreta, lo podrá hacer también en otras tareas en contextos diferentes. Un ejemplo de este enfoque es el examen de elementos discretos donde aspectos como el léxico se evalúa independientemente (Purpura, 2016).

En el enfoque que se basa en tareas (*the task-based approach*), el *construct* es todo el desempeño de una tarea. Esta tarea debe ser definida en cuanto al escenario (puede ser una farmacia u otro lugar del mundo real) o, por ejemplo, el contexto comunicativo (es decir, el

significado del mensaje que queremos transmitir). Si el estudiante ha logrado completar la tarea, habitualmente se juzga según los mismos criterios que definen la realización exitosa de la tarea en la vida real. El tipo de evaluación que ha sido desarrollado gracias a este enfoque es la evaluación basada en tareas (Purpura, 2016), que será discutida en los siguientes capítulos.

1.1.4 La fiabilidad y el impacto de la evaluación

Otro término que se relaciona con el desarrollo y la organización de una evaluación es la fiabilidad. Según Bachman y Palmer (1996), hace referencia a la consistencia de los resultados. Si el estudiante hiciera el examen otra vez bajo las mismas condiciones (sin aprender nada de la previa prueba), debería obtener el mismo resultado para que el examen tenga alta fiabilidad. Los mismos autores usan el término «impacto» para hablar de las consecuencias que una evaluación puede tener. Hamps-Lyons (2009) usa *washback* como sinónimo para el impacto y afirma que puede ser positivo o negativo (Stoynoff, 2009). Lo último significa que el profesor modifica el contenido de las clases para preparar a los estudiantes para una evaluación cuyo problema es que no refleja el uso real de la lengua (Abrar-ul-Hassan et al., 2021).

1.1.5 La ética de la evaluación

Como la evaluación puede tener consecuencias graves para los estudiantes, es necesario tener en cuenta su aspecto ético. La comunicación entre el profesor y el estudiante sobre las pruebas, la puntuación y su interpretación debería ser transparente. Toda evaluación tiene que ser justa para no desmotivar al estudiante del aprendizaje. Deberíamos usar varios tipos de evaluación porque cada estudiante es diferente y trabajar con un solo tipo nunca puede revelar todo el conocimiento de todos (Azizi et al., 2023).

1.2 La evaluación a gran escala

1.2.1 Definición y *maturita*

La evaluación a gran escala no es producto de los profesores, sino que la desarrolla un participante externo al aula. Se usa para examinar a los estudiantes en varias escuelas y habitualmente se hace a través de pruebas estandarizadas que nos dejan fácilmente comparar los resultados. Suele ser un examen de alto riesgo que sirve como entrada a otro tipo de educación o como certificado de lo que ya el estudiante ha logrado. La estadística obtenida así se puede revelar al público o aprovechar para comparar los resultados de escuelas diferentes con el fin de señalar a los responsables de la calidad de la educación cuál escuela necesita más ayuda (McKay, 2005), por ejemplo, financiera.

El examen a gran escala relevante para el presente estudio es *maturita*. En la República Checa es un examen al final del instituto y abarca varias asignaturas. La parte llamada «común» contiene las asignaturas definidas por la ley llamada «la ley escolar»¹. Todos los estudiantes tienen que hacer un examen de checo, y pueden elegir entre matemáticas y una lengua extranjera. Los exámenes tienen forma de prueba escrita distribuida a todos los estudiantes al mismo tiempo y evaluada por autoridades fuera de la escuela concreta (Zkoušky a společné části, 2019).

Si el estudiante elige la prueba de una lengua extranjera en la parte común, tiene que hacer el examen de la misma lengua también en la parte de *maturita* diseñada por la escuela (esta parte corresponde al tipo de escuela y su currículo) (Centrum pro zjišťování výsledků vzdělávání, 2019). Sin embargo, también puede elegir hacer el examen, aunque en la parte

¹ (Zkoušky a společné části, 2019) 561/2004 (Právní předpisy, 2019)

común no haya optado por la lengua extranjera, sino por matemáticas. Aquí la prueba contiene la redacción de un texto y una prueba oral (Centrum pro zjišťování výsledků vzdělávání, 2019). La escuela se encarga de la evaluación de esta parte (Centrum pro zjišťování výsledků vzdělávání, 2019).

1.2.2 Beneficios

Un gran beneficio de la evaluación a gran escala es su imparcialidad, ya que normalmente hay otra persona que evalúa el examen en lugar del profesor del estudiante concreto. Otros beneficios son la fácil organización (habitualmente el examen tiene formato de papel) y la eficiencia financiera (gracias al hecho de que se distribuye a muchos estudiantes al mismo tiempo) (McKay, 2005).

1.2.3 La influencia en la enseñanza

A veces las creencias de los profesores sobre qué deberían enseñar y las creencias de los creadores de los exámenes a gran escala difieren. Sin embargo, como este tipo de evaluación es normalmente de alto riesgo, los profesores no deberían subestimar su importancia para los estudiantes (National Research Council, 2003). Por eso, una desventaja de esta evaluación es el fenómeno llamado «enseñar para los exámenes». Significa que el profesor da prioridad a la preparación para el examen a gran escala, es decir, se enfoca en los temas que aparecerán en el examen y, por eso, reduce el currículo o entrena a los estudiantes en el uso de estrategias útiles para aprobar el examen. El problema con esta actitud es que cada evaluación a gran escala es limitada y no puede abarcar todo el conocimiento de un tema presentado en la clase. Entonces, alinear el contenido de las clases con el examen puede hacer que omitamos una parte significativa del currículo. Además, los exámenes a gran escala casi siempre dejan de medir habilidades complejas como la capacidad resolutoria o la cooperación. Tienden a emplear

estrategias sencillas, como ensayos cortos o preguntas de opción múltiple, para abarcar un gran número de estudiantes (Looney, 2009).

1.2.4 Límites de la evaluación a gran escala

Una importante pregunta es si estos formatos o los exámenes a gran escala en general pueden realmente captar las habilidades lingüísticas de los estudiantes. Por ejemplo, hay pruebas que se dividen en secciones, con una dedicada al uso del léxico y otra al uso de la gramática. Sin embargo, esto no refleja cómo la lengua se usa en la vida real fuera del proceso de evaluación (McKay, 2005).

Un examen que no refleja qué es un idioma y cuáles son las competencias lingüísticas puede causar mucho daño. Gulfranz Abbasi et al. (2010) investigan el impacto de la evaluación a gran escala en el aprendizaje de lenguas en Pakistán. El examen usado en el país no incluye producción oral ni comprensión auditiva y, debido al contenido de la prueba, una estrategia exitosa para aprobarla es la memorización. Como esta evaluación es bastante importante, casi todos los profesores alinean el currículo de sus clases con ella. Los autores describen que la mayoría de los estudiantes que participaron en su estudio usaron la memorización para aprender la lengua y no pueden comunicarse o entender audios. Este es un ejemplo de cómo un examen a gran escala puede perjudicar a los estudiantes y el sistema educativo.

1.3 La evaluación en clase

1.3.1 Definición de la evaluación en clase

Quienes preparan esta evaluación y la llevan a cabo en el contexto del aprendizaje en clase son los profesores, aunque no todos los exámenes hechos en el aula son ejemplos de evaluación en clase. Por ejemplo, la *maturita* checa también se realiza en el aula y con la participación del profesor, pero no se considera parte de la evaluación en clase porque ha sido preparada por otra autoridad jerárquicamente superior (McKay, 2005).

La evaluación puede ser sumativa (medición del progreso del estudiante tras un determinado período de tiempo) o formativa (investigamos cuáles son los puntos fuertes y débiles del estudiante para modificar el proceso de aprendizaje y cubrir sus necesidades). Por eso, la importancia de la evaluación en clase es doble: queremos ayudar a los estudiantes y evaluar su progreso. También puede ser planeada o espontánea. Si es planeada, el profesor puede pensar en términos objetivos, el *construct* a evaluar, tareas y estrategias que se usarán, propósitos de los resultados, el umbral de éxito y oportunidades de mejora. Si es espontánea, ocurre durante la clase cuando el profesor lo cree necesario, y no es planeada (McKay, 2005).

La evaluación en clase habitualmente se complementa con la evaluación a gran escala en situaciones evaluativas de alto riesgo, como la admisión al instituto o a la universidad. La razón es que los padres pueden dudar de la validez y fiabilidad de la evaluación en clase y desear que una autoridad superior se encargue (McKay, 2005).

1.3.2 Los *constructs*

Yin (2010) escribe sobre los *constructs* que los profesores tienen en cuenta cuando observan a los estudiantes (habitualmente para la evaluación formativa). Estos *constructs* se dividen en tres grupos: los relacionados con la lengua, los relacionados con el plan de estudios o los relacionados con el profesor. Los primeros son, por ejemplo, la pronunciación, el léxico o la gramática. Los segundos son habilidades mencionados en el programa de estudios, como hacer una presentación. Los que se relacionan con el profesor son los *constructs* que él considera importantes a pesar de que no estén en el plan de estudios o no se relacionen con la lengua. Un ejemplo de esto puede ser el pensamiento crítico cuando la tarea para los estudiantes es descubrir falacias en un texto (Yin, 2010).

1.3.4 Las creencias de los profesores

Los profesores a veces creen en estereotipos, es decir, expectativas generales de lo que el estudiante puede o no puede hacer. Se basan en experiencias previas con estudiantes parecidos; por ejemplo, de la misma nacionalidad o trabajo (*au-pairs*, médicos, empresarios). Estos estereotipos a veces influyen en la interpretación de los resultados de la evaluación y en cómo el profesor enseña. Las creencias sobre la gente fuera del aula también pueden modificar el proceso de aprendizaje. Los profesores ocasionalmente proporcionan una retroalimentación basada en cómo ellos imaginan que un hablante nativo reaccionaría cuando oyera la producción oral del estudiante (Yin, 2010).

No obstante, tener «una imagen mental» del estudiante no es siempre perjudicial. A veces, los profesores las crean como parte de la evaluación de clase mientras las clases se están desarrollando. Por ejemplo, pueden identificar el nivel de la competencia lingüística del estudiante cuando lo oyen hablar o cuando se fijan en que interacciona rápidamente y con fluidez. Esta información también influye en las decisiones tomadas, particularmente cuando el

profesor divide a los estudiantes en grupos según sus competencias o el nivel de ayuda y atención necesitada. Un problema con «las imágenes mentales» es que son impresionistas, es decir, que no nos brindan un resumen completo y sistemático de lo que el estudiante puede hacer y necesitar. Por lo tanto, es aconsejable complementarlas con una evaluación más sistemática. Entender todos los elementos de cómo los profesores piensan cuando hacen una evaluación es clave para mejorar la evaluación misma (Yin, 2010).

McMillian (2003) investiga qué influye en cómo los profesores toman decisiones en cuanto a la evaluación. Sostiene que se basan en sus creencias, pero también consideran las consecuencias para los estudiantes (por ejemplo, tener confianza en sí mismos o un sentido de logro). También piensan en los objetivos establecidos por el currículo. Pueden usar una variedad de evaluaciones para responder a las diferencias entre los estudiantes, ya que evaluaciones distintas nos pueden mostrar capacidades diferentes de los estudiantes (McKay, 2005).

Sus creencias pueden estar en conflicto con los requerimientos de la escuela donde trabajan (Crusan et al., 2016). La consecuencia de esto puede ser que haya diferencia entre lo que se debería evaluar y lo que realmente se evalúa. Asimismo, si las autoridades superiores requieren el uso de una evaluación específica, a veces algunos profesores o, incluso las escuelas, lo rechazan en silencio y siguen usando lo que corresponde a sus creencias (Yin, 2010).

1.3.6 Diferencias entre la evaluación a gran escala y la evaluación en clase

Shepard (2013) menciona varias diferencias entre la evaluación a gran escala y la evaluación en clase. La de a gran escala necesita ser estandarizada, pero la hecha en clase puede ser dinámica y adaptarse al estudiante y a lo que necesitamos evaluar en el momento. La de a gran escala se hace en un día determinado, pero la hecha en clase se realiza cuando hay oportunidad para ella durante el aprendizaje en clase. La última también ofrece espacio para

que el profesor ayude con el desempeño (por ejemplo, la evaluación dinámica que será mencionada en los capítulos siguientes). Un beneficio de la evaluación en clase es que el profesor puede proporcionar retroalimentación. Esto normalmente no sucede cuando los estudiantes hacen un examen a gran escala. La precisión técnica de un examen a gran escala es más estricta que la de la evaluación en clase porque es más de alto riesgo (National Research Council, 2003).

Para resumir las diferencias, podemos decir que la evaluación a gran escala es conveniente cuando el sistema educativo quiere datos sumativos, cuantitativos y de un gran número de estudiantes. Estos datos luego se pueden usar para justificar el gasto público y las políticas de educación. La evaluación en clase es útil cuando el profesor necesita información individualizada sobre el progreso de un estudiante con el fin de modificar el proceso de aprendizaje (Davis, 2018).

1.4 La evaluación sumativa

Este tipo de evaluación se hace al final de un periodo de aprendizaje (por ejemplo, al final de un capítulo en el libro de texto o al final del año escolar). El objetivo es resumir el progreso que el estudiante ha hecho. Este resumen (habitualmente en forma de puntuación) se puede usar para tomar decisiones sobre el futuro aprendizaje del estudiante (por ejemplo, si podrá ingresar en la universidad). La evaluación sumativa no consta de informaciones diagnósticas sobre los puntos débiles y fuertes del estudiante o no nos dice cómo podemos modificar el proceso de estudio para ayudar a la persona. El enfoque es averiguar si el estudiante ha hecho suficiente progreso (Ismail et al., 2022).

Los métodos de esta evaluación son pruebas estandarizadas que a menudo resultan en una nota (Muñoz Restrepo et al., 2020). Como es fácil distribuir estas pruebas, se usan en muchos países como exámenes a gran escala para medir si los estudiantes han logrado las metas del aprendizaje (Ismail et al., 2022).

Como un examen sumativo es a veces de alto riesgo, puede causar estrés en el estudiante y es posible que el resultado nos dé más informaciones sobre si la persona es capaz de trabajar bajo presión que sobre su aprendizaje. La evaluación se lleva a cabo en un momento determinado y por eso no nos brinda datos sobre el desempeño del estudiante durante un período más largo. Esto también significa que si usamos solo la evaluación sumativa, habrá períodos en el proceso de aprendizaje donde la evaluación no ocurre. Esto puede causar que el estudiante se sienta desmotivado durante este tiempo y estudie solo cuando el examen sumativo se acerca. Sin embargo, este problema no se puede solucionar con la substitución de la evaluación de alto riesgo con la de bajo riesgo. Si introducimos una serie de exámenes de bajo riesgo en el aula, la ausencia de consecuencias graves puede también contribuir a la desmotivación de los

estudiantes (Korpi, 2019). Además, debido a los exámenes tan a menudo presentes, los estudiantes pueden tener la impresión de que todo el aprendizaje consta de una evaluación (Banerjee et al., 2016).

1.5 La evaluación formativa

1.5.1 Introducción y definición

Uno de los papeles del profesor es supervisar el progreso del estudiante y tomar decisiones frente a la trayectoria del aprendizaje. La evaluación formativa, que facilita esta labor (Burke, 2002), consiste en detectar lo que el estudiante (no) ha aprendido (Burke, 2002) y, así, dar pasos de cara al futuro proceso de aprendizaje (Xiao et al., 2019) y cubrir las necesidades del estudiante (Vassiliou et al., 2022). La evaluación formativa es un proceso continuo que se puede desempeñar durante una variedad de actividades (Xiao et al., 2019); por ejemplo, observaciones, discusiones durante las clases, evaluación entre pares y autoevaluación o retroalimentación proporcionada por el profesor (Vassiliou et al., 2022).

Como el objetivo es mejorar el proceso de aprendizaje, a veces la evaluación formativa se llama «evaluación para el aprendizaje» (*assessment for learning*). Por su parte, la evaluación sumativa se describe como «evaluación del aprendizaje» (*assessment of learning*), ya que busca resumir lo que se logra al final del proceso de aprendizaje. La evaluación para el aprendizaje implica dar información a los estudiantes sobre el progreso que han hecho junto con consejos sobre lo que debe mejorarse (Vassiliou et al., 2022).

La diferencia entre las dos evaluaciones no es siempre clara. A veces no podemos decir que un tipo de evaluación es siempre formativo o sumativo, sino que depende de su propósito. Además, es importante destacar que la evaluación formativa no sirve para reemplazar la sumativa en todos los casos. Idealmente, las dos deberían complementarse según nuestros objetivos (Vassiliou et al., 2022). También se pueden mezclar las dos. La sumativa se puede

hacer más formativa cuando después del examen el estudiante no solo recibe la puntuación, sino también puede revisar el examen y obtener retroalimentación (Ismail et al., 2022).

La decisión de usar una u otra evaluación en ocasiones depende del riesgo que el examen presenta. Más riesgo significa que los evaluadores probablemente optarán por la sumativa. Esto ocurre porque los exámenes a gran escala pueden representar la entrada a un programa educativo y el evaluador quiere determinar si el candidato tiene las competencias suficientes. En otras palabras, el objetivo es resumir lo que la persona ha aprendido (Tsagari et al., 2016).

Una parte importante de la evaluación formativa es la retroalimentación. Debería ser alentadora y crear confianza en los estudiantes mismos (Burke, 2002). También se necesita espacio para que el estudiante, en diálogo con el profesor, entienda la retroalimentación que recibe (Xiao et al., 2019).

Hay un vínculo importante entre el MCER y la evaluación formativa. Una de las metas del MCER es controlar el aprendizaje (en relación con sus descriptores) y promover una actitud positiva hacia la educación a través del uso de los descriptores *can-do*; así se refleja lo que los estudiantes ya pueden hacer con sus habilidades lingüísticas. El Consejo de Europa valora la evaluación formativa como una actitud clave para lograr esta meta (De Neve et al., 2022).

1.5.2 La evaluación formativa en el aula

Los profesores se pueden sentir vacilantes frente a la tarea de introducir esta evaluación en el aula. Idealmente, deberían tener acceso, por ejemplo, a recursos de instrucción, materiales, ejemplos de las estrategias de aprendizaje. Una fuente importante de apoyo puede ser un grupo de profesores con quienes pueden discutir sus logros y sus luchas. Asimismo, observar a sus colegas mientras usan la evaluación puede convencerlos de que el cambio es valioso y real. Según Andersson y Palm (2018), es más probable que los profesores implementen cambios en

sus aulas si creen que estos cambios serán exitosos y que son importantes y útiles (De Neve et al., 2022).

Hay un *toolkit* en línea (formatieevaluieren.kdg.be) con métodos e ideas de la evaluación formativa. Fue creado tras un estudio sobre los procesos de aprendizaje de los profesores que habían introducido esta evaluación en sus clases. Contiene ejemplos de cómo observar a los estudiantes y analizar su trabajo, cómo involucrarlos en las discusiones sobre la retroalimentación que reciben y cómo guiarlos hacia la autoevaluación. El propósito del *toolkit* es brindar consejos útiles y una base de datos con metodología a los profesores que quieren implementar la evaluación formativa. Esto puede ayudar a los profesores vacilantes para que se sientan más seguros durante la introducción de los cambios (De Neve et al., 2022).

El estudio de De Neve et al. (2022) también investiga la introducción de la evaluación formativa en cuatro escuelas secundarias en Bélgica. Todas estas escuelas tenían un número siempre creciente de estudiantes de comunidades diversas y querían promover actividades lingüísticas. Tres de las cuatro escuelas eran privadas en un entorno urbano y estaban financiadas con fondos públicos. En el estudio, que se llevó a cabo entre 2017 y 2019, tenían los profesores el *toolkit* mencionado anteriormente. Este *toolkit* les dio el marco teórico para la evaluación formativa junto con ejemplos de actividades que pudieran usar (De Neve et al., 2022).

Los que ya habían usado la evaluación formativa regularmente en sus clases aseveraron que tener acceso al *toolkit* les ayudó a ser más sistemáticos y conscientes en lo que hacían. Muchos profesores mencionaron que la evaluación formativa requería mucho tiempo. Algunos elementos no se podían usar en todos los contextos; por ejemplo, había dificultades con la evaluación entre pares en cuanto a los errores cometidos por los compañeros. Los profesores que tenían la posibilidad de colaborar con sus colegas y solían aprovechar su tiempo para

discutir los valores críticos para la enseñanza o la necesidad de usar la evaluación formativa sacaron provecho del apoyo mutuo (De Neve et al., 2022).

1.5.3 Beneficios

La evaluación formativa está alienada con las tendencias educativas en los países de la Organización para la Cooperación y el Desarrollo Económicos (OECD), que promueven el aprendizaje y la autonomía de los estudiantes (Looney, 2009). Apoyan la auto-regulación de los estudiantes, es decir, la habilidad de identificar sus propios objetivos de aprendizaje, probar diferentes estrategias para estudiar, evaluarse a sí mismos y cooperar con los otros para obtener retroalimentación. Por lo tanto, los estudiantes pueden desarrollar su interés por el aprendizaje y ser activos en la búsqueda de estrategias para superar las dificultades académicas que enfrentan. Según Nicol y Macfarlane-Dick (2006), algunos elementos de la evaluación formativa que promueven la auto-regulación incluyen dejar al estudiante participar en el establecimiento de los objetivos, escoger cómo cumplirlos o determinar cómo responder a la retroalimentación. Los profesores deberían también formular criterios claros para la evaluación e invitar a los estudiantes a usarlos para evaluar su propio trabajo. Los estudiantes también deberían tener la oportunidad de evaluar a sus compañeros (Xiao et al., 2019). Si se involucran en la creación de criterios transparentes, aprenden a evaluar y a autoevaluar (Burke, 2002).

Cuando los estudiantes tienen la oportunidad de pensar sobre cómo y cuánto aprenden y tienen los recursos necesarios para mejorar, aumenta la probabilidad de que sus habilidades y conocimientos crezcan (Davis, 2018). También, su motivación puede ser más alta si los profesores les proporcionan evidencia tangible de lo que ya han aprendido. Como la evaluación formativa habitualmente no es de alto riesgo, los estudiantes experimentan menos ansiedad causada por la evaluación (Burke, 2002).

1.5.4 Problemas con la evaluación formativa

Ya hemos mencionado que esta forma de evaluar en ocasiones aumenta la carga para los profesores (De Neve et al., 2022). Otro problema es su continuidad. Se da a lo largo de las clases y es necesario, aunque difícil, que el profesor sea consistente con la información que obtiene, aunque se trate de un tiempo largo. Gatullo (2002) estudia el uso de la evaluación formativa en las escuelas italianas. En su estudio descubrió que los profesores tienden a recoger información, pero a no usarla después. Tampoco aprovechan todas las situaciones donde pueden indagar algo importante. Por ejemplo, omiten preguntas metacognitivas, es decir, dejar que los estudiantes expliquen su manera de pensar. Esto significa que no obtiene toda la información importante sobre el proceso de aprendizaje de sus estudiantes (McKay, 2005). La evaluación formativa puede ser exigente a causa de la atención prolongada que requiere y el aumento del trabajo que significa para los profesores.

1.5.5 Otras formas de la evaluación formativa

En los siguientes capítulos se mencionarán otras estrategias de evaluación formativa.

1.6 Evaluación basada en tareas

1.6.1 Introducción y definición

Desde los años 60 del siglo XX, se ha desarrollado un nuevo enfoque en la enseñanza de idiomas. Este enfoque denominado *Communicative Language Teaching* presenta el estudio de idiomas no como el aprendizaje de reglas lingüísticas, sino como la adquisición de herramientas para la comunicación en situaciones reales (Giraldo, 2020).

De conformidad con este desarrollo nació *task-based language teaching* (TBLT) y después *task-based language assessment* (TBLA) durante las últimas décadas del siglo XX (Noroozi, 2021). En TBLT, los estudiantes aprenden un idioma desempeñando tareas auténticas y significativas (Davis, 2018). TBLA entonces mide las habilidades del estudiante al momento de usar el idioma para completar las tareas. El enfoque está en el desempeño y no en los conocimientos lingüísticos abstractos (Noroozi, 2021) porque el idioma es visto como una herramienta para sobrevivir en situaciones sociales. El mero conocimiento de la gramática y del léxico no es suficiente para comunicarse con éxito en contextos reales. El usuario debe saber también cómo usar la lengua una manera efectiva para conseguir sus metas comunicativas (Mislevy et al., 2002).

1.6.2 Enfoques en TBLA

De acuerdo con Mislevy et al. (2002), hay dos enfoques en TBLA. Los profesores pueden evaluar si los estudiantes han conseguido las metas de la tarea (este enfoque se llama *task-centred assessment* y se fija en qué tareas suscitan las habilidades que queremos evaluar). O pueden evaluar aspectos diferentes del desempeño como la gramática o el léxico (por ejemplo, hasta qué punto el estudiante ha usado una gama del léxico específico) (este enfoque

se llama *construct-centred assessment*). El segundo enfoque nos brinda informaciones sobre las habilidades que el estudiante puede usar fuera de la situación evaluada). Una tarea puede consistir en dar direcciones para ir a una biblioteca. Podemos evaluar si el estudiante ha logrado explicar cómo llegar a la biblioteca (*task-centred assessment*) o nos podemos enfocar en elementos diferentes, como la complejidad del mensaje transmitido, el contenido, la precisión de la gramática usada, el léxico o la pragmática, es decir, si el desempeño fue adecuado en el contexto sociolingüístico (*construct-centred assessment*). Se debe tener en cuenta que los hablantes a veces economizan el idioma durante situaciones comunicativas y, por eso, no todos los elementos están siempre presentes al máximo punto de las habilidades del estudiante. Por este motivo los profesores deberían determinar los componentes clave para el cumplimiento de la tarea y prestar atención a estos (Mislevy et al., 2002). Según Bachman (2002), los dos enfoques se deberían usar simultáneamente. Prestar atención solo al hecho de que la meta de la tarea (no) ha sido conseguida puede resultar problemático porque el éxito de una tarea en el proceso de aprendizaje de lenguas extranjeras no significa únicamente que el estudiante sea capaz de desempeñar la tarea en contextos diferentes (Giraldo, 2020).

1.6.3 Beneficios de TBLA

TBLA incluye objetivos claros, es decir, el estudiante tiene en cuenta las metas que debe conseguir. También se promueve el uso de la lengua auténtica, la retroacción formativa y otros tipos de la evaluación alternativa como evaluación entre pares (*peer-assessment*). Se desarrollan habilidades comunicativas útiles en la vida real (Giraldo, 2020). TBLA puede ser útil en clases con estudiantes jóvenes porque ellos aprenden a través de actividades concretas y significativas (McKay, 2005).

1.6.4 Limitaciones de TBLA

Ellis (2003) y Skehan (1984) identifican como dificultades de TBLA el problema de autenticidad, representatividad, generalizabilidad e inseparabilidad. Aunque las tareas deberían imitar situaciones de la vida real, no es posible conseguir autenticidad perfecta. La situación presentada y la lengua usada deberían copiar la vida real y los participantes auténticos deberían estar involucrados. Sin embargo, es muy difícil imitar estas condiciones en el entorno del aula. Además, podemos sostener que los exámenes son siempre artificiales y, por eso, no podemos llegar al nivel de autenticidad deseado (Giraldo, 2020).

El significado de generalizabilidad es doble. Puede referirse a cuán consistente es el desempeño del estudiante a través de tareas diferentes, o puede significar el poder predictivo de la tarea en TBLA – si el estudiante cumple la tarea con éxito, sobre qué cantidad de situaciones reales podemos predecir que en ellas el estudiante también será exitoso. Si la relación entre la tarea en TBLA y la situación de la vida real no está claramente definida, la generalizabilidad es baja (Noroozi, 2021).

1.7 La evaluación dinámica

1.7.1 Vygotsky y la Zona de Desarrollo Próximo

La evaluación dinámica como una alternativa a los métodos tradicionales surgió gracias a la teoría sociocultural del psicólogo ruso L. S. Vygotsky y del científico israelí Reuven Feuerstein. Según esta teoría, el aprendizaje ocurre a través de la mediación, es decir, a través del contacto con los demás, quienes nos proporcionan retroalimentación, corrigen nuestros errores y nos presentan nuevas ideas. De esta manera, podemos modificar nuestro comportamiento, incluso los diferentes aspectos de la lengua que usamos, y ampliar nuestras habilidades (Poehner et al., 2021).

Una parte importante de esta teoría es la Zona de Desarrollo Próximo (*the Zone of Proximal Development* o ZDP). Este término se refiere a las habilidades que no somos capaces de demostrar solos, sino con ayuda de los demás (Poehner et al., 2021) (o con ayuda de artefactos culturales (Vergara Cabarcas et al., 2022), como los diccionarios). Vygotsky (1978) define la ZDP como la distancia entre el nivel del desarrollo donde podemos solucionar problemas independientemente y el nivel de desarrollo potencial donde podemos solucionar problemas con ayuda de alguien más hábil (García, 2019). La colaboración con alguien más experto nos ayuda a integrar paso a paso lo que todavía no sabemos usar solos (Poehner et al., 2021). Por eso, Vygotsky sostiene que la ZDP es el lugar donde ocurre el aprendizaje (Antón, 2009). En el aula, enfocarnos en la ZDP nos sirve para saber a dónde dirigir el proceso de enseñanza-aprendizaje para ayudar al estudiante (García, 2019).

Es importante resaltar que la ZDP no se puede definir como una habilidad y que tampoco podemos hablar de una ZDP más o menos grande; es una metáfora de cómo ocurre el aprendizaje, no un dato concreto (García, 2019). Kozulin y Garb (2002) acuñaron el término «potencial de aprendizaje» (*learning potential*), que puede sustituir la ZDP (Poehner et al., 2021).

1.7.2 Beneficios y limitaciones de la evaluación dinámica

La información que obtenemos a través de la evaluación dinámica nos sirve como diagnóstico del proceso de aprendizaje. Podemos detectar qué componentes del currículum representan un problema para el estudiante. Asimismo, este enfoque promueve el aprendizaje porque las pistas y demás ayuda que el estudiante recibe le indican cómo mejorar y desarrollar sus habilidades. Cuando vemos lo que el estudiante es capaz de hacer con apoyo, podemos predecir su futura trayectoria de aprendizaje. En contraste, el enfoque tradicional – la evaluación sumativa – nos brinda solo información sobre el desempeño actual. También podemos decir que la evaluación dinámica se enfoca más en el proceso de adquirir la lengua, mientras que la evaluación sumativa se concentra en el producto del aprendizaje que ha ocurrido (Vergara Cabarcas et al., 2022).

Este tipo de evaluación es especialmente útil cuando tenemos que dividir estudiantes en grupos según el nivel de sus competencias. Si usamos la evaluación sumativa, obtenemos resultados sobre sus previas experiencias en el sistema educativo. Esto puede ser perjudicial para los estudiantes que no han tenido acceso a una educación de buena calidad. En cambio, con el uso de la evaluación dinámica podemos encontrar formas para promover sus potenciales y captar una imagen más compleja de lo que pueden hacer (Yeld et al., 1997).

En el entorno checo, esto es relevante, por ejemplo, en el primer año en el instituto, cuando los estudiantes se dividen en grupos con base en sus niveles de inglés. Por lo general,

vienen de una variedad de escuelas que ofrecen clases de inglés de diferente calidad. Usar un examen estandarizado y sumativo nos revelaría información sobre estas escuelas y no necesariamente sobre el potencial de los estudiantes. Tampoco indicaría si el estudiante puede aprender rápidamente con la ayuda adecuada o si necesita estudiar con un grupo donde los conceptos nuevos se enseñen de forma menos exigente.

A pesar de sus beneficios, la evaluación dinámica no ha sido adoptada ampliamente. La razón puede ser que requiera mucho tiempo debido a que se hace individualmente con cada estudiante y, por eso, los profesores optan por otros tipos de evaluación. Otra desventaja es que los resultados no se convierten fácilmente en puntuación o evidencias tangibles de las habilidades de los estudiantes (Antón, 2009). Entonces este tipo de evaluación sigue siendo una alternativa poco usada.

1.8 La autoevaluación

1.8.1 Característica de la autoevaluación

La evaluación formativa hace énfasis en que los estudiantes tomen parte activa en el proceso (Davis, 2018). La autoevaluación es un ejemplo de evaluación formativa donde los estudiantes participan activamente del proceso (Geeslin, 2003).

Para ayudar a los estudiantes a entender los criterios según los cuales deberían evaluar su trabajo, el profesor les puede proporcionar rúbricas evaluativas en donde se describen los componentes clave del desempeño de una tarea. Esto conduce a los estudiantes a reflexionar sobre las diferentes partes del desempeño, a fijarse en lo que han hecho con éxito y en lo que necesitan mejorar. Si queremos asegurarnos de que la rúbrica sea objetiva, podemos incluir elementos como el número de errores que el estudiante ha cometido o el número de revisiones necesarias para que el estudiante cumpla la tarea (Geeslin, 2003).

Pierce et al. (1993) sostienen que es más probable que los estudiantes y los profesores lleguen al mismo resultado de la evaluación si los criterios se especifican claramente; por eso, la rúbrica debe detallarlos. Por ejemplo, debería describirse un comportamiento específico para que el estudiante compare un aspecto de su desempeño con esta descripción. Algunas partes de la rúbrica pueden también servir como impulsos para que el estudiante compare su comportamiento con el de sus compañeros. Por ejemplo, las descripciones pueden decir «Levanté mi mano más que los demás.» o «En mi grupo, yo soy el líder.». A la última se puede añadir aún más detalle para que quede claro lo que se evalúa. La descripción, por lo tanto, puede ser: «Soy el líder del grupo porque siempre empiezo a trabajar en la tarea lo más pronto

posible.» o «Soy el líder del grupo porque, cuando alguien necesita ayuda, siempre le ayudo.» (Geeslin, 2003).

1.8.2 Beneficios de la autoevaluación

No hay evidencia de que mostrar una rúbrica evaluativa a los estudiantes mejora su desempeño en, por ejemplo, la producción escrita. Sin embargo, Becker (2016) encontró que si los estudiantes participan en el diseño de la rúbrica, tienen mejor entendimiento de los criterios de evaluación y de las expectativas del proceso formativo. Esto los conduce a mejorar su competencia escrita. Una limitante en este estudio es que se enfoca solo en un tipo de producción escrita (el resumen) en un grupo pequeño de estudiantes.

Además, la autoevaluación promueve la confianza de los estudiantes en sí mismos (Geeslin, 2003). La rúbrica evaluativa puede también aumentar la coherencia de la autoevaluación cuando comparamos evaluadores diferentes, es decir, si la misma rúbrica ha sido usada, es más fácil comparar la autoevaluación de dos estudiantes diferentes (Becker, 2016).

1.8.3 Limitaciones de la autoevaluación

A pesar de que la autoevaluación tiene muchos beneficios, debería complementarse con otros tipos de evaluación para que no pongamos toda la responsabilidad en el juicio del estudiante. Algunos tienden a sobrestimarse, otros a subestimarse (Cárdenas Ramos, 1997). Esto se ha observado, por ejemplo, en la autoevaluación de chicas (Erickson et al., 2022).

Usar la autoevaluación en la clase con niños puede ser problemático porque su experiencia con la evaluación puede ser limitada y es posible que no entiendan los criterios. Muchos libros de texto incluyen *can-do statements*, es decir, oraciones sobre lo que el estudiante puede hacer con el fin de que el estudiante elija si es verdad para él o no. Sin embargo, los niños a veces entienden de otro modo las palabras usadas en estas oraciones (como

los verbos como «seguir» o «entender»). El problema se intensifica con el hecho de que muchos descriptores se usan generalmente en muchos libros de texto y no son especializados para cada grupo de estudiantes o cada individuo. Son vagos y pueden ser malinterpretados. Por ejemplo, la oración «Puedo entender la conversación de otras personas.» se puede entender como «Entiendo las conversaciones que tenemos en la clase» o «Entiendo la conversación en la lengua extranjera si oigo a alguien hablar fuera de la clase» (Davis, 2018).

1.9 La evaluación entre pares

1.9.1 Definición

La evaluación entre pares es similar a la autoevaluación. También la hace el estudiante, pero esta vez no evalúa su propio trabajo, sino el trabajo de su compañero. Si queremos usar estas dos evaluaciones en el aula, necesitamos entrenar a los estudiantes en cómo hacerlas correctamente. Por ejemplo, para la evaluación entre pares los estudiantes tienen que saber cómo proporcionar retroalimentación a los demás sin menospreciar su trabajo y desmotivarlos (McKay, 2005).

1.9.2 Beneficios

Durante la evaluación entre pares, los estudiantes aprenden a ayudar a sus compañeros y a permitir que un tercero los ayude. Por eso, se puede decir que tanto la evaluación entre pares como la autoevaluación no son solo una estrategia de evaluación, sino también una estrategia de enseñanza. La primera también contribuye a que el estudiante se sienta miembro de una comunidad (McKay, 2005).

1.9.3 Limitaciones

En ocasiones a los estudiantes no les gusta usar este tipo de evaluación (por ejemplo, cuando deben producir notas), a pesar de que tienden a ser menos estrictos con el fin de ayudar a sus compañeros y animarlos (Sridharan et al., 2017). No obstante, los estudiantes no suelen confiar en la evaluación producida por sus compañeros; para mitigar esta desconfianza la evaluación hecha de esta manera debería ser detallada y específica. Si es vaga, causa frustración en los evaluados. Además, la retroalimentación debería incluir consejos de cómo mejorar y

ejemplos de respuestas correctas. También debería ser personalizada y respetuosa. Si los estudiantes piensan que al evaluador no le importa su trabajo y dar retroalimentación, se enfadan y no creen en la evaluación. Tampoco ayuda si no tienen la oportunidad de discutir la evaluación con el evaluador. Este tipo de evaluación siempre debería ser de bajo riesgo para todos los participantes, porque la relación entre el evaluado y el evaluador es complicada y el evaluador no es un profesional (Zhou et al., 2020).

1.10 Portafolios

1.10.1 Definición de portafolio

El portafolio sirve como una colección del trabajo del estudiante. Es evidencia de su progreso tras un período de tiempo y puede ser evaluado. Los trabajos incluidos en la colección deberían seguir instrucciones claras sobre qué tipo de evidencia se debería recoger y los criterios de evaluación. Lo que se incluye puede ser diverso; por ejemplo, dibujos, ensayos u otros textos, obras de arte o autoevaluaciones. El portafolio es una buena oportunidad para la autorreflexión y también presenta una base para las discusiones con el profesor sobre el progreso del estudiante (McKay, 2005).

El portafolio no se usa tanto para la evaluación sumativa, como para la evaluación formativa. Cárdenas Ramos (1997), por ejemplo, afirma que un portafolio evaluativo debe incluir los comentarios del profesor sobre el trabajo del estudiante. Sostiene que también el estudiante debería incluir sus comentarios sobre el profesor y su enseñanza.

1.10.2 Beneficios y limitaciones

En los portafolios, los estudiantes pueden demostrar su mejor trabajo. También el uso de este tipo de evaluación disminuye el estrés del estudiante porque no supone necesariamente la comparación con el trabajo de los compañeros. Los portafolios son también ideales para mostrar el progreso del estudiante durante las reuniones padres-maestros (Cárdenas Ramos, 1997).

Un problema de los portfolios es que exige tiempo porque el profesor debe evaluar toda la colección de cada estudiante. También se encuentra el problema de subjetividad porque, si

las instrucciones lo permiten, uno puede incluir solo su mejor trabajo en el portafolio (Cárdenas Ramos, 1997). También lo que los profesores piensan sobre un portafolio es a veces subjetivo. Esto significa dificultades con la fiabilidad, es decir, diferentes profesores pueden evaluar el mismo portafolio de diferentes modos. También surge otro problema con el tiempo, pues hay que darle suficiente tiempo al estudiante para que genere muchos ejemplos de lo que puede hacer (McKay, 2005).

2 Metodología

2.1 El estudio

El objetivo de este trabajo es averiguar si los profesores y las profesoras en las escuelas checas usan los tipos de evaluación mencionados anteriormente. Se investigan sus actitudes hacia la evaluación y las razones por las cuales usan o no usan ciertos tipos. El tipo de la investigación es cualitativo porque el interés son las opiniones de los profesores.

Las hipótesis son tres: 1. Cuando los profesores eran estudiantes, sus profesores no usaban la evaluación formativa, sino la sumativa. 2. Los profesores (participantes del estudio) combinan la evaluación formativa con la sumativa, pero la formativa la limitan a la retroalimentación. Es decir, no usan la evaluación dinámica, basada en tareas, portafolios, la autoevaluación o la evaluación entre pares. 3. A la hora de evaluar, los profesores se enfrentan con presión por parte de los padres y las escuelas porque ellos requieren que los resultados de los estudiantes sean buenos.

2.2 Participantes

Como muestra la Tabla 1, todos los participantes del presente estudio, cuya lengua materna es el checo, son profesores de *gymnázium* (un tipo de escuela secundaria en Chequia). En cuanto a su formación profesional, tres profesores tienen formación pedagógica. En cuanto a la duración de la experiencia profesional, tres de los profesores ya han enseñado más de quince años, uno seis años y uno está empezando su carrera profesional y no ha trabajado en la escuela ni un año escolar completo.

Tabla 1

Datos básicos sobre los participantes

Profesor	Nacionalidad	Lengua materna	Formación pedagógica	Años de experiencia profesional	Tipo de escuela
1	checa	checo	no, formación filológica + <i>pedagogické minimum</i> ²	6	<i>osmileté gymnázium</i> ³
2	checa	checo	grado pedagógico, máster filológico + <i>pedagogické minimum</i> requerido por la escuela en el futuro	0,6	<i>čtyřleté⁴ y osmileté gymnázium</i>
3	checa	checo	sí	más de 15	<i>čtyřleté y osmileté gymnázium</i>
4	checa	checo	sí	más de 15	<i>čtyřleté y osmileté gymnázium</i>
5	checa	checo	sí	más de 15	<i>čtyřleté y osmileté gymnázium</i>

² *Pedagogické minimum* no existe como un término en la ley de trabajadores pedagógicos (esta ley habla de formación o cursos adicionales) (*Zákon o pedagogických pracovnících a o změně některých zákonů*, 2004), pero se usa para referir a esta formación adicional pedagógica que los sin formación universitaria de este tipo pueden obtener para poder trabajar como profesores.

³ *osmileté gymnázium* - un tipo de escuela que ofrece currículo desde el sexto curso hasta *maturita*

⁴ *čtyřleté gymnázium* – un tipo de escuela que ofrece currículo desde el curso siguiente al noveno curso hasta *maturita*

2.3 Instrumentos

Para obtener la información, se desarrolló una serie de preguntas abiertas (ver Anexo 1) que se enfocaron en caracterizar al profesor (por ejemplo, en cuanto al tipo de escuela donde enseña o cuántos años de experiencia profesional tiene) y en determinar qué tipos de evaluación usa y por qué. Esta segunda parte sobre diferentes tipos de evaluación contenía 24 preguntas divididas en secciones y cada sección correspondía a un capítulo en la parte teórica. Sin embargo, las preguntas no contenían los nombres de los tipos de evaluación por dos razones: primero, en caso de que los profesores no conocieran los nombres, enfrentarse con un texto lleno de ellos los podría disuadir de participar en el estudio; segundo, era posible que no separaran qué elementos de lo que hacían para evaluar pertenecerían qué tipo. Si el estudio les preguntara si usaban la evaluación formativa, podrían, por equivocación, omitir estos elementos por falta de conocimientos teóricos sobre esta evaluación.

2.4 Procedimiento

Primero se hizo una versión piloto para comprobar que todas las preguntas eran claras para los encuestados potenciales y para asegurarse de que no faltaran oportunidades para obtener informaciones clave para el estudio. Para llevar a cabo esta versión, se hizo una videollamada con un participante piloto quien también tiene experiencia profesional como profesor. Respondió a las preguntas para que quedara claro si el instrumento cubría todos los temas que este trabajo investiga.

Después de la versión piloto, 13 profesores y profesoras de 4 escuelas diferentes recibieron una invitación para participar en el estudio junto con las preguntas para que pudieran tener una visión clara de lo que trataba la investigación. 6 profesores no respondieron a la invitación, otros 7 expresaron su deseo de participar. De estos 7, solo 5 acordaron hacer una cita para discutir las preguntas. 3 citas se hicieron presencialmente, 2 a través de una

videollamada. Durante toda la entrevista pudieron también mirar las preguntas y había espacio para que modificaran sus respuestas durante la entrevista o inmediatamente después de ella. Todas las respuestas fueron apuntadas por la autora del estudio durante la entrevista en una hoja con las preguntas impresas.

2.5 Análisis de datos

Como se trata de un estudio cualitativo y la mayoría de los datos proviene de preguntas abiertas, el análisis consta de comparar las respuestas de los participantes para encontrar rasgos comunes de las respuestas e identificar las tendencias en la práctica y los pensamientos de los profesores.

3. Resultados

3.1 Experiencia general con la evaluación

Todos los profesores afirman que durante sus estudios hacían exámenes en forma escrita y recibían notas. Dos de ellos también obtenían comentarios sobre sus resultados. Durante sus estudios pedagógicos o el proceso de obtener *pedagogické minimum* generalmente no había muchas oportunidades para expandir sus conocimientos sobre la evaluación. El contenido de los estudios abordó el tema muy brevemente y de manera teórica. Una persona menciona estudiar la diferencia entre el tipo sumativo y el tipo formativo. También es el único participante que usa la palabra «sumativa» para hablar sobre lo que pasa durante los procesos de evaluación en sus clases. Otros ocasionalmente mencionan la evaluación formativa, pero a veces no están seguros de lo que significa y lo admiten.

Dos profesores no han obtenido formación adicional sobre cómo evaluar después de sus estudios universitarios. Tres (aquellos con más años de experiencia profesional) han aprendido más del tema en seminarios especializados y mencionan que también les ha ayudado mucho el

proceso de enseñar y observar lo que funciona o discutir con los colegas. Una persona destaca la importancia de *maturita* para la propagación de la alfabetización. Como el examen contiene descriptores para que los profesores puedan evaluar la producción escrita de los estudiantes, los profesores pueden observar las reglas de la evaluación y, según el participante, incluirlas en sus aulas.

Tres profesores admiten que necesitan más formación en el tema de evaluación. Mencionan el deseo de participar en seminarios y hablar con sus colegas sobre sus experiencias o tener la oportunidad de aclarar los aspectos de la evaluación donde no se sienten seguros (por ejemplo, si evaluar el contenido o el hecho de que el estudiante transmita el mensaje que desea). Una persona destaca la importancia de tener conocimientos de problemas psicosociales para trabajar con las generaciones contemporáneas y aboga por incluir estos temas en la formación de los profesores, incluso cuando se trata del proceso de evaluar.

3.2 Evaluación a gran escala: *maturita*

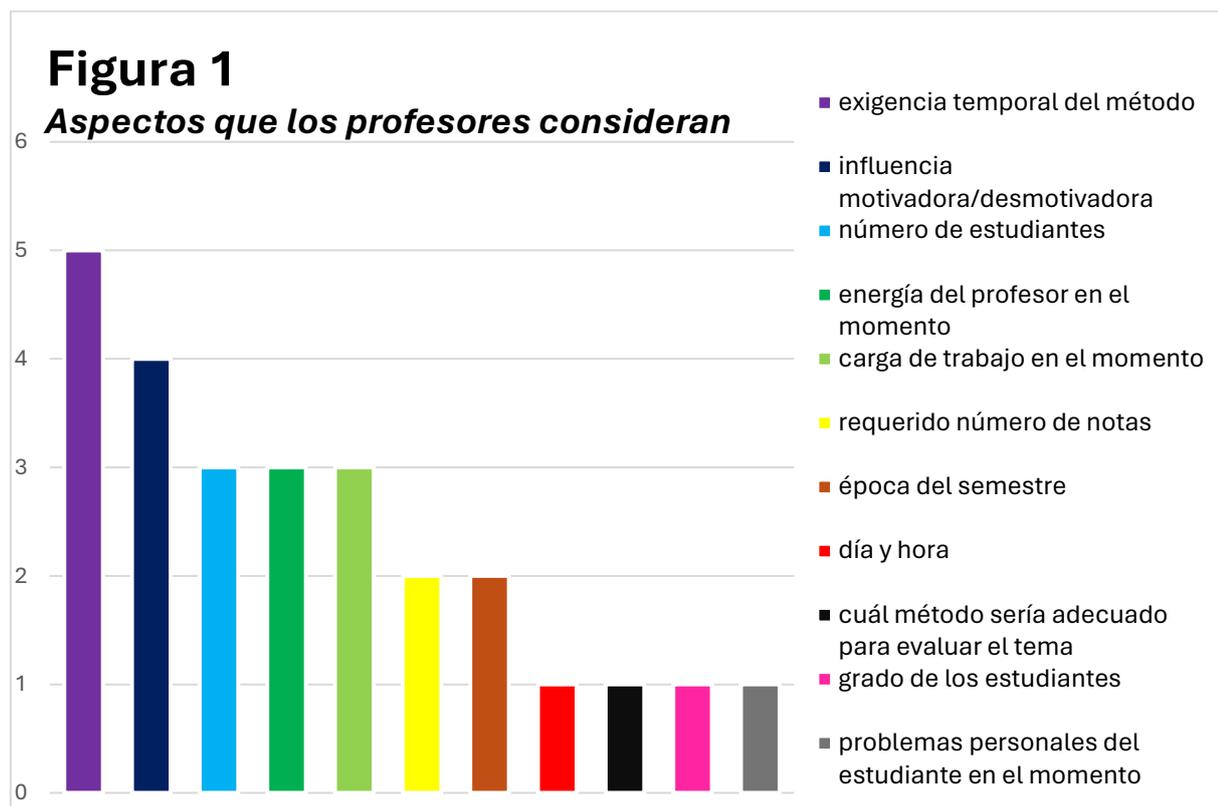
Todos los profesores están de acuerdo en que tener *maturita* como un examen final es bueno porque es necesario concluir de manera oficial los estudios en el instituto. No obstante, los profesores critican el contenido del examen. Uno menciona que la parte organizada por las escuelas contiene preguntas sobre la cultura en los países donde se habla la lengua, y otros temas, a pesar de que esto no esté directamente relacionado con el conocimiento de la lengua o con los descriptores de las habilidades en el MCER. A otro participante le parece que *maturita* en esta forma no es útil porque habitualmente no sirve como examen de admisión a las universidades. En términos generales, los profesores no están en contra de *maturita*, pero perciben fallas de contenido y de su rol en el aprendizaje de lenguas.

Cuatro profesores enseñan clases en los dos grados previos a *maturita*. Todos dicen que el examen influye en lo que enseñan. Las opiniones sobre si su influencia es positiva o negativa

difieren. El participante que lo ve como negativo critica que la preparación para el examen no deja mucho tiempo para enseñar otros temas del currículo. Al contrario, el participante que considera dicha influencia extremadamente positiva destaca la ventaja de extender el contenido de las clases para incluir aspectos del examen que no aparecen en el libro del estudiante, y así proporcionar más conocimientos a los estudiantes. Aunque las opiniones difieren, podemos resumir el tema diciendo que *maturita* influye en las clases.

3.3 Evaluación en clase

Hay varios elementos que los profesores toman en consideración cuando evalúan a los estudiantes. La Figura 2 resume esta información.



Todos los profesores tienen en cuenta la exigencia temporal del método usado para la evaluación. Cuatro consideran la influencia motivadora o desmotivadora para los estudiantes y tres también piensan en su energía, el número de estudiantes evaluados y su carga de trabajo. Por ejemplo, un profesor trabaja y estudia en la universidad a la vez y por eso incluye solo tales

actividades evaluativas en las clases que para él no exigen mucho trabajo fuera del aula. Otro profesor prefiere métodos que producen materiales físicos, como los exámenes escritos, para tener evidencia tangible de los resultados frente a las quejas de los padres y los estudiantes.

Dos profesores mencionan sentirse un poco bajo la presión de los padres a la hora de la evaluación, pero afirman que al final no influye en el resultado. Nadie percibe presión por parte de los estudiantes. Hay ciertas influencias que causan las escuelas o el sistema; por ejemplo, tener que otorgar evaluación en forma de comentarios escritos al final del año escolar, usar notas o usar una escala de porcentajes y las notas correspondientes creada por la escuela.

Los profesores admiten que el proceso de evaluación difiere entre ellos porque cada uno se enfoca en aspectos diferentes y les da niveles distintos de importancia; afirman que sería muy complicado si sus colegas tuvieran que evaluar a sus estudiantes. Por eso, aceptarían esta situación solo como un experimento. Algunos reconocen que sus colegas necesitarían saber al menos el contenido de las clases para poder evaluar a los estudiantes. Una persona le también proporcionaría informaciones sobre su objetivo, los requerimientos y a qué aspectos presta atención cuando evalúa. Un participante teme que los colegas menos empáticos hieran a sus estudiantes y por eso no les dejaría evaluarlos.

3.4 Evaluación formativa

Todos los profesores proporcionan retroalimentación a sus estudiantes y lo ven como una parte importante de la evaluación. Comentan los resultados de los exámenes y el desempeño de los estudiantes en la clase. Cuatro profesores comprueban el progreso de los estudiantes. De estos cuatro, dos lo hacen sin planearlo y habitualmente durante las actividades en las clases escuchando a los estudiantes. Uno usa el libro de trabajo con *can-do statements* y uno usa una aplicación en su móvil donde sube el léxico que los estudiantes deberían estudiar y luego en la

aplicación puede ver el progreso que los estudiantes han hecho. Todos los profesores proporcionan consejos a los estudiantes de cómo mejorar su desempeño.

Usar la evaluación entre pares no es popular entre los participantes del estudio. Les parece que los estudiantes pueden decir algo negativo que dañaría a sus compañeros de clase o tratarían de «ayudar» a los demás y les darían las mejores notas. Por eso, si a veces usan este tipo de evaluación, por ejemplo, cuando los estudiantes hacen presentaciones, no tiene efecto en los resultados oficiales.

Tres profesores tampoco usan la autoevaluación. Uno usa los *can-do statements* que ya hemos mencionado y uno deja espacio en los exámenes escritos para que los estudiantes evalúen al profesor y se evalúen a sí mismos. Pero dice que casi nadie lo utiliza porque los estudiantes no saben cómo autoevaluarse porque esto no forma parte de la cultura checa.

Asimismo, nadie usa los portafolios. Los argumentos para esto son que los estudiantes y los profesores ya tienen demasiado trabajo, que los cuadernos de los estudiantes se pueden usar como portafolios, y que los exámenes escritos no se pueden incluir en una compilación de esta manera porque los profesores necesitan guardarlos como evidencia.

3.5 Evaluación sumativa

Todos los profesores incluyen exámenes escritos sumativos en la evaluación y otorgan notas, generalmente porque el sistema escolar lo requiere y porque los estudiantes están acostumbrados a recibir notas, y eso los motiva. Dos profesores afirman que prefieren no tener notas y sustituirlas con el sistema *pass/fail* durante el año escolar o con porcentajes para no estigmatizar al estudiante y poder expresar su progreso con más detalle.

3.6 Evaluación basada en tareas

La mayoría de los profesores (4 de 5) emplea actividades que imitan la vida real en sus clases; por ejemplo, pedir comida en el restaurante o discutir con sus compañeros. Casi todos se enfocan en si la actividad se completa con éxito y no en los componentes individuales como léxico o gramática. Uno destaca que el objetivo es mostrarles a los estudiantes que pueden comunicarse en otra lengua (y, así, subir su autoestima).

3.7 Evaluación dinámica

En cuanto a los errores cometidos por los estudiantes, 3 profesores los corrigen solo si les parece un error grave durante el desempeño de la actividad. Por ejemplo, no interrumpen al estudiante mientras está discutiendo con los demás si comete un error que no afecta negativamente el entendimiento del mensaje.

Si los estudiantes cometen un error que merece corrección, los profesores (al menos 3 de ellos) les advierten que han cometido un error pero los dejan descubrirlo independientemente y cambiar lo que han dicho o escrito. Las actitudes hacia la puntuación en este caso difieren. Un profesor otorga puntos como si el estudiante no cometiera un error si logra corregirlo. Otro da solo la mitad de los puntos para reflejar que el desempeño no es perfecto al principio. Un profesor destaca que el sistema escolar checo castiga los errores y que es perjudicial y desmotivador para los estudiantes. Dice que obtener una nota mala no es tan grave para el estudiante como sentirse castigado por algo que es natural en el proceso de aprendizaje.

3.8 Discusión

La primera hipótesis de esta tesis era que los profesores experimentaban solo la evaluación sumativa cuando ellos mismos eran estudiantes. La investigación muestra que esto no es verdad para todos los participantes, ya que dos también obtenían retroalimentación. Es

decir, la idea de la evaluación formativa ya estaba presente en el sistema educativo. Sin embargo, como podemos ver, si comparamos estos recuerdos de los participantes con lo que ellos mismos emplean en sus clases como profesores, el desarrollo de métodos nuevos ha sido más prominente en las últimas décadas. Aunque los profesores no usan términos como «la evaluación basada en tareas» o «la evaluación entre pares», mencionan muchas actividades que conducen a la evaluación formativa.

No obstante, hay una tendencia significativa frente a las actividades que eligen dentro de todo el rango formativo. Optan por actividades que no cuestan mucho tiempo (por eso rechazan, por ejemplo, el portafolio) o no representan una carga mental para los estudiantes, quienes han crecido en el sistema escolar checo y les costaría acostumbrarse a una evaluación distinta de lo que suelen recibir. Los hallazgos de la investigación están de acuerdo con De Neve et al. (2022) y Korpi (2019), quienes afirman que la evaluación formativa a veces aumenta la carga de trabajo de los profesores. Esto confirma la segunda hipótesis en la que se proponía que los profesores, cuando usan la evaluación formativa, prefieren la retroalimentación (que es fácil de organizar y no cuesta tanto tiempo). Pero hay que destacar la presencia de otros métodos que la retroalimentación como, por ejemplo, las tareas que reflejan la vida real o indicar al estudiante que ha cometido un error sin ofrecer la solución correcta inmediatamente.

Los profesores mencionan que, para ampliar sus conocimientos, necesitan el apoyo de sus colegas. Esta afirmación también está de acuerdo con el estudio de De Neve et al. (2022), que menciona la importancia de tener una red de colegas con quienes puedan discutir sobre sus dudas y retos para que los profesores introduzcan cambios en la evaluación más fácilmente.

Los participantes de este estudio combinan la evaluación formativa con la sumativa. Es verdad que el sistema educativo checo la prefiere y requiere que los profesores usen las notas. Sin embargo, parece que los profesores aceptan esta situación y, si tienen objeciones, se trata

de la forma (introducir el sistema *pass/fail* en vez de las notas) y no la evaluación sumativa como tal.

La tercera hipótesis sugiere que, a la hora de evaluar, los profesores se enfrentan con la presión de evaluar causada por los padres y las escuelas, debido a que quieren que los resultados de los estudiantes sean buenos. La investigación muestra que esto puede ser falso. Los profesores niegan sentirse bajo presión de los padres o las escuelas. La única presión que les importa es la presión de usar las notas. Pero es interesante fijarse en los otros aspectos de su trabajo que influyen en cómo evalúan. Entre otros, destacan la exigencia temporal del método usado para evaluar y la influencia motivadora y desmotivadora para los estudiantes. Esto indica que los profesores tienen que economizar con el tiempo en su trabajo, pero también piensan en los estudiantes y el impacto que la evaluación tiene.

Los hallazgos sobre *maturita* están de acuerdo con otros estudios sobre la evaluación a gran escala. Según Looney (2009), los profesores habitualmente son conscientes del alto riesgo que representa un examen a gran escala y por eso «enseñan para el examen», es decir, en vez de enseñar los materiales que normalmente ven como importantes, dan prioridad a la preparación de los estudiantes para la prueba. La investigación confirma que los profesores que trabajan con clases uno o dos años antes del examen final experimentan este fenómeno.

En conclusión, los profesores emplean un rango de evaluaciones y ven la evaluación formativa como una parte importante del proceso de aprendizaje. Optan por métodos que les parecen útiles y posibles de realizar tomando en cuenta su carga de trabajo, el tiempo y la actitud de los estudiantes, y las costumbres en el sistema escolar.

3.9 Limitaciones del estudio y vías de investigación

Aunque el estudio trata de investigar, entre otras cosas, qué tipos de evaluación los profesores usan, es importante notar que todos los datos se basan en las respuestas proporcionadas por los profesores y, por eso, dependen de lo que ellos admitieron durante la entrevista y de su imagen de cómo enseñan. Los datos nos dicen más sobre lo que ellos creen que hacen y no lo que realmente hacen. Otra limitación surge del tipo de la escuela donde los profesores trabajan. A pesar de que se trata de escuelas diferentes, todos son *gymnázium*, cuyo objetivo es preparar a los estudiantes para los estudios en una universidad. Si bien la intención era incluir también profesores de *základní škola* (escuela primaria), no fue posible debido a su falta de interés en la participación en el estudio. El número de participantes del estudio es también muy bajo para que podamos generalizar los resultados sobre el sistema educativo checo. Se necesitaría hacer un estudio con más participantes de diferentes tipos de escuelas. También sería bueno incluir encuestados de nacionalidades y lenguas maternas diferentes, ya que todos los participantes son checos y no solamente profesores checos enseñan español en el país. Un próximo estudio debería también incluir un grupo más variado según el número de años de experiencia profesional.

4 Conclusión

El objetivo de esta tesis era investigar qué tipos de evaluación usan los profesores checos. De las tres hipótesis, el trabajo confirma dos, pero solo parcialmente. Es verdad que los profesores combinan la evaluación sumativa y formativa y que de la última generalmente prefieren dar retroalimentación, pero a diferencia de la hipótesis, usan también otras formas de la formativa.

La tercera hipótesis, que los profesores se enfrentan a presión por parte de los padres y las escuelas a la hora de evaluar, no se confirmó. La única presión importante proviene del sistema educativo, que tradicionalmente prefiere la evaluación sumativa y el uso de notas.

El resto del estudio brinda información detallada sobre los tipos de evaluación que los profesores usan y por qué. Ya se ha establecido que usan la sumativa y la formativa. De la última, prefieren la retroalimentación y tareas que reflejan la vida real.

Como los participantes del estudio forman un grupo bastante homogéneo en cuanto a sus nacionalidades y las escuelas donde los profesores trabajan (todos son checos y enseñan *gymnázium*), se necesitaría darle continuidad a este estudio con participantes de diferentes sistemas educativos (en cuanto a su formación, como a su carrera profesional). Las futuras investigaciones podrían incluir un estudio de las actitudes hacia la evaluación de profesores trabajando en diferentes tipos de escuelas.

Anexos

Anexo 1

CUESTIONARIO - EVALUACIÓN EN LA ENSEÑANZA DE LENGUAS EXTRANJERAS EN LAS ESCUELAS CHECAS

Informaciones básicas

Nacionalidad:

Lengua materna:

Educación superior pedagógica sí no en:

Tipo de escuela:

Años de experiencia profesional con la enseñanza de idiomas:

menos de 2 años

2-5 años

5-10 años

10-15 años

más de quince años

Si quiere impartir el número preciso:



¿Qué tipos de evaluación usaban sus profesores en la enseñanza de lenguas extranjeras?

¿Era el tema de evaluación parte de sus estudios pedagógicos?

¿Ha recibido alguna formación en cómo evaluar después de estos estudios?

¿Quisiera tener más formación en los procesos de evaluación? Cree usted que necesita más? ¿Por qué sí/no?

¿Qué necesitaría para tener aún mejor entendimiento de este tema?



¿Maturita es una manera adecuada de evaluación al final del instituto

¿Maturita influye en cómo enseña? (la selección de materiales, la práctica del examen con los estudiantes) (“cambiar la manera de enseñar”)

Si es así, ¿considera esta influencia positiva o negativa?



Cuando evalúa el trabajo de los estudiantes, ¿toma en consideración

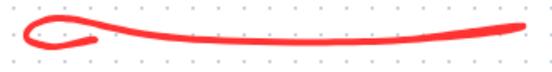
la exigencia temporal del método el número de estudiantes la influencia motivadora/desmotivadora para los estudiantes

el requerido número de notas el día y la hora la energía de usted en el momento la época del semestre carga de trabajo en el momento

otro aspecto?

¿Se siente forzado por parte de los padres/la escuela/los estudiantes a evaluar de alguna manera?

¿Dejaría a otro profesor/otra profesora que evaluara a sus estudiantes al final de un ciclo de aprendizaje? Si es así, ¿qué tipo de información le proporcionaría antes de la evaluación?



¿Proporciona retroalimentación a los estudiantes durante el proceso de aprendizaje (aparte de las pruebas clasificadas)? Si es así, ¿en qué forma y por qué?

¿Comprueba el progreso de los estudiantes durante el proceso de aprendizaje (aparte de las pruebas clasificadas)? Si es así, ¿de qué manera?

¿Deja a los estudiantes que evalúen el trabajo de los demás? Si es así, ¿qué importancia tiene esta evaluación (por ejemplo, si influye en la nota que el estudiante recibe)?

¿Deja a los estudiantes que se autoevalúen? Si es así, ¿qué importancia tiene esta evaluación (por ejemplo, si influye en la nota que el estudiante recibe)?

¿Deja/invita a los estudiantes que crean su portafolio? Si es así, ¿qué debería estar en el portafolio?

¿Habla con los estudiantes sobre su progreso y les da recomendaciones sobre cómo mejorar?



¿En sus clases hay pruebas que resuman lo que los estudiantes han aprendido?

¿Da notas a los estudiantes? ¿Por qué sí/no?



¿Entrena con los estudiantes cómo comunicarse en situaciones que correspondan con actividades reales (como pedir comida en un restaurante, discutir en un grupo...)?

¿Evalúa el desempeño de estas tareas? Si es así, ¿qué es lo que evalúa (el cumplimiento exitoso de la tarea, la gramática correcta, el uso del léxico...)?



¿Si un estudiante comete un error, lo corrige?

¿Es el estudiante siempre independiente en el proceso de evaluación o puede recibir ayuda?

Si es así, ¿hay casos donde el estudiante y usted trabajan como equipo para la evaluación?

Referencias

Abbasi, M. G., Ahmad, A., & Khattak, Z. I. (2010). Negative influence of large scale assessment on language learning strategies of the Secondary School Certificate (SSC) students. *Procedia, social and behavioral sciences*, 2(2), 4938-4942. <https://doi.org/10.1016/j.sbspro.2010.03.799>

Abrar-ul-Hassan, S., Douglas, D., & Turner, J. (2021). Revisiting second language portfolio assessment in a new age. *System (Linköping)*, 103, 102652. <https://doi.org/10.1016/j.system.2021.102652>

Ahmed Abdel-Al Ibrahim, K., Karimi, A. R., Abdelrasheed, N. S. G., & Shatalebi, V. (2023). Group dynamic assessment vs. computerized dynamic assessment: impacts on L2 listening development, L2 students' perfectionism, foreign language anxiety, and intrinsic motivation. *Language Testing in Asia*, 13(1), 32-29. <https://doi.org/10.1186/s40468-023-00245-1>

Al-Hawamdeh, B. O. S., Hussen, N., & Abdelrasheed, N. S. G. (2023). Portfolio vs. summative assessment: impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13(1), 12-29. <https://doi.org/10.1186/s40468-023-00225-5>

Andujar, A. (2020). Mobile-Mediated Dynamic Assessment: A New Perspective for Second Language Development. *ReCALL (Cambridge, England)*, 32(2), 178-194. <https://doi.org/10.1017/S0958344019000247>

Anton, M. (2009). Dynamic Assessment of Advanced Second Language Learners. *Foreign language annals*, 42(3), 576-598. <https://doi.org/10.1111/j.1944-9720.2009.01030.x>

Assessment in support of instruction and learning bridging the gap between large-scale and classroom assessment - workshop report. (c2003). National Academies Press.

Azizi, Z., & Ismail, S. M. (2023). What ethical requirements should be considered in language classroom assessment? insights from high school students. *Language Testing in Asia*, 13(1), 23-22. <https://doi.org/10.1186/s40468-023-00235-3>

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language testing*, 19(4), 453-476. <https://doi.org/10.1191/0265532202lt240oa>

Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing writing*, 29, 15-24. <https://doi.org/10.1016/j.asw.2016.05.002>

Burke, R. (2002). Formative Assessment Procedures and the Second Language Curriculum: Signposts for the Journey. *TESL Canada journal*, 19(2), 87. <https://doi.org/10.18806/tesl.v19i2.931>

Cárdenas Ramos, R. (1997). Exploring Possibilities for Alternative Assessment in Foreign Language Learning. *Íkala: revista de lenguaje y cultura*, 2(1-2), 57-73. <https://doi.org/10.17533/udea.ikala.8044>

Centrum pro zjišťování výsledků vzdělávání (2019). “Maturitní zkouška.” <https://maturita.ceremat.cz/menu/maturitni-zkouska>

Centrum pro zjišťování výsledků vzdělávání (2019). “Právní předpisy.” <https://maturita.ceremat.cz/archiv-maturitni-zkousky/mz2022-archiv/pravni-predpisy-2022-archiv>

Centrum pro zjišťování výsledků vzdělávání (2019). “Zkoušky profilové části.” <https://maturita.ceremat.cz/menu/maturitni-zkouska/zkousky-profilove-casti>

Centrum pro zjišťování výsledků vzdělávání (2019). “Zkoušky společné části.” <https://maturita.ceremat.cz/menu/maturitni-zkouska/zkousky-spolecne-casti>

Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing writing*, 28, 43-56. <https://doi.org/10.1016/j.asw.2016.03.001>

Cummins, P. W., & Davesne, C. (2009). Using Electronic Portfolios for Second Language Assessment. *The Modern language journal (Boulder, Colo.)*, 93(s1), 848-867. <https://doi.org/10.1111/j.1540-4781.2009.00977.x>

De Neve, D., Leroy, A., Struyven, K., & Smits, T. (2022). Supporting formative assessment in the second language classroom: an action research study in secondary education. *Educational action research*, 30(5), 828-849. <https://doi.org/10.1080/09650792.2020.1828120>

Erickson, G., Borger, L., & Olsson, E. (2022). National assessment of foreign languages in Sweden: A multifaceted and collaborative venture. *Language testing*, 39(3), 474-493. <https://doi.org/10.1177/02655322221075067>

Exploring the Role of Dynamic Assessment in Language Education: An Introduction to the Special Issue. (2019). *Language and sociocultural theory*. <https://doi.org/10.1558/1st.38912>

Geeslin, K. L. (2003). Student Self-Assessment in the Foreign Language Classroom: The Place of Authentic Assessment Instruments in the Spanish Language Classroom. *Hispania*, 86(4), 857-868. <https://doi.org/10.2307/20062958>

Georgetown University Round Table on Languages and Linguistics (2016: Washington, D. C.), & Davis, J. M. E. (2018). *Useful Assessment and Evaluation in Language Education*. Georgetown University Press.

Giraldo, F. (2020). Task-Based Language Assessment: Implications for the Language Classroom. *GIST Education and Learning Research Journal*, 21(21), 209-224. <https://doi.org/10.26817/16925777.828>

Hamp-Lyons, L. (2009). Principles for Large-Scale Classroom-Based Teacher Assessment of English Learners' Language: An Initial Framework From School-Based Assessment in Hong Kong. *TESOL quarterly*, 43(3), 524-530. <https://doi.org/10.1002/j.1545-7249.2009.tb00249.x>

Ismail, S. M., Rahul, D. R., Patra, I., & Rezvani, E. (2022). Formative vs. summative assessment: impacts on academic motivation, attitude toward learning, test anxiety, and self-regulation skill. *Language Testing in Asia*, 12(1), 1-23. <https://doi.org/10.1186/s40468-022-00191-4>

Korpi, S. (2019). Portfolio Project as Summative Language Assessment: Engaging Learners Online. *International journal of e-learning & distance education*, 34(2), 1-18.

- Looney, J. (2009). Assessment and Innovation in Education. <https://doi.org/10.1787/222814543073>
- McKay, P. (2005). Assessing language use through tasks. In *Assessing Young Language Learners* (pp. 97–139). chapter, Cambridge: Cambridge University Press.
- McKay, P. (2005). Classroom assessment of language use. In *Assessing Young Language Learners* (pp. 140–175). chapter, Cambridge: Cambridge University Press.
- McKay, P. (2005). Research into the assessment of young language learners. In *Assessing Young Language Learners* (pp. 61–96). chapter, Cambridge: Cambridge University Press.
- Mckay, P. (2005). Testing young language learners through large-scale tests. In *Assessing Young Language Learners* (pp. 315-351). Cambridge University Press. <https://doi.org/10.1017/CBO9780511733093.010>
- Ministerstvo školství, mládeže a tělovýchovy (2023). “Zákon o pedagogických pracovnících a změně některých zákonů.” <https://www.msmt.cz/dokumenty/zakon-o-pedagogickych-pracovnicich-ve-zneni-ucinnem-ode-dne>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing*, 19(4), 477-496. <https://doi.org/10.1191/0265532202lt241oa>
- Muñoz Restrepo, A., Valderrama Carvajal, I., Lopez Muñoz, A., & Avendaño Franco, R. (2020). Measuring Language Gains in a Foreign Language Context. *Colombian applied linguistics journal*, 22(1), 69-86. <https://doi.org/10.14483/22487085.14171>
- Noroozi, M. (2021). A Survey of the Issue of Generalizability in Task-based Language Assessment. *Theory and practice in language studies*, 11(8), 884-890. <https://doi.org/10.17507/tpls.1108.03>
- Poehner, M. E., & Infante, P. (2016). Dynamic Assessment in the language classroom. In *Handbook of Second Language Assessment* (Vol. 12, pp. 275-290). De Gruyter. <https://doi.org/10.1515/9781614513827-019>
- Poehner, M. E., & Wang, Z. (2021). Dynamic Assessment and second language development. *Language teaching*, 54(4), 472-490. <https://doi.org/10.1017/S0261444820000555>
- Purpura, J. E. (2016). Second and Foreign Language Assessment. *The Modern language journal (Boulder, Colo.)*, 100(S1), 190-208. <https://doi.org/10.1111/modl.12308>
- Stoynoff, S. (2009). Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language teaching*, 42(1), 1-40. <https://doi.org/10.1017/S0261444808005399>
- Takala, S., Erickson, G., Figueras, N., & Gustafsson, J. -eric. (2016). Future Prospects and Challenges in Language Assessments. In *Contemporary Second Language Assessment* (pp. 299-316). Bloomsbury Publishing. <https://doi.org/10.5040/9781474295055.ch-015>
- Tsagari, D., & Banerjee, J. (2016). The handbook of second language assessment. In *Handbook of Second Language Assessment* (Vol. 12, pp. 1-10). De Gruyter. <https://doi.org/10.1515/9781614513827-003>
- Tsagari, D., & Banerjee, J. V. (2016). Linking to the CEFR: Validation Using a Priori and a Posteriori Evidence. In *Contemporary Second Language Assessment*. Bloomsbury Publishing.

Turnbull, B. (2020). Towards new standards in foreign language assessment: learning from bilingual education. *International journal of bilingual education and bilingualism*, 23(4), 488-498. <https://doi.org/10.1080/13670050.2017.1375891>

Vassiliou, S., Papadima-Sophocleous, S., & Giannikas, C. N. (2022). *Formative Assessment in Second Language Learning: A Systematic Review and an Annotated Bibliography*. Research-publishing.

Vergara Cabarcas, L. K., Castellon Barrios, D. J., López Caraballo, J. L., Vásquez Rossi, C. A., & Becker Arroyo, E. A. (2022). Dynamic Assessment Approach in Language Teaching: A Review. *Zona próxima*, (30), 82-99. <https://doi.org/10.14482/zp.30.371.3>

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System (Linköping)*, 81, 39-49. <https://doi.org/10.1016/j.system.2019.01.004>

Yeld, N., & Haecck, W. (1997). Educational Histories and Academic Potential: can tests deliver? *Assessment and evaluation in higher education*, 22(1), 5-16. <https://doi.org/10.1080/0260293970220101>

Yin, M. (2010). Understanding Classroom Language Assessment Through Teacher Thinking Research. *Language assessment quarterly*, 7(2), 175-194. <https://doi.org/10.1080/15434300903447736>

Zhou, J., Zheng, Y., & Tai, J. H. -M. (2020). Grudges and gratitude: the social-affective impacts of peer assessment. *Assessment and evaluation in higher education*, 45(3), 345-358. <https://doi.org/10.1080/02602938.2019.1643449>