

# Proposal:

Item Response Theory and Confirmatory Factor  
Analysis equivalence: Application on a teacher  
evaluation process in Peru

**Jose Manuel RIVERA ESPEJO**

Supervisor: Prof. A. Xyz  
*Affiliation (optional)*

Co-supervisor: *(optional)*  
*Affiliation (optional)*

Mentor: *(optional)*  
*Affiliation (optional)*

Proposal presented in  
fulfillment of the requirements  
for the degree of Master of Science  
in Statistics and Data Science:  
Social, Behavioral and Educational Sciences

Academic year 2020-2021

# 1 Topic justification and context

order:

- Paragraph's main point: Why it is essential to identify good teachers

Hincapié et al (2020) - Profesores a prueba se deben incluir políticas que abarquen los diferentes aspectos que determinan dicha calidad. Por el lado de la oferta educativa, sin duda, una de las variables fundamentales es la calidad de los docentes.

Throughout the literature much has been emphasized about the positive impacts of the teachers in learning ability of the students

- Paragraph's idea 1: impact on student learning

Bertoni et al (2020) - Concursos Docentes en Latinoamérica De hecho, tener un buen docente tiene un efecto importante en disminuir las brechas de aprendizajes (Araujo et al., 2016; Chetty et al., 2014; Hanushek y Rivkin, 2012; Rockoff, 2004), sobre todo en las escuelas que atienden a estudiantes de menor rendimiento (Rivkin, Hanushek y Kain, 2005),

Bertoni et al (2020) - Escasez de docentes en Latinoamérica La evidencia disponible sugiere que los maestros temporales aplican mayores niveles de esfuerzo y pueden tener un impacto positivo en el rendimiento de los estudiantes, cuando la renovación del contrato depende de su desempeño (Duflo et al., 2009; Muralidharan y Sundararaman, 2013).

- Paragraph's idea 2: impact on teacher composition
- Paragraph's idea 3: impact beyond the classroom and in the long run

Bertoni et al (2020) - Concursos Docentes en Latinoamérica e incluso repercute en resultados de largo plazo como decisiones de educación superior, desempeño laboral o aspectos sociales, tales como el embarazo adolescente (Chetty Friedman y Rockoff, 2014b).

- Paragraph's idea 4: what happens when you don't have good teachers

Bertoni et al (2020) - Escasez de docentes en Latinoamérica los docentes contratados por esta vía tienen atributos inferiores a aquellos que ganan las vacantes. Esto es muy relevante, especialmente para aquellos países que están elevando los estándares mínimos para la contratación permanente

... Sin embargo, cuando los contratos temporales no están sujetos a rendición de cuentas, existe evidencia de que los maestros temporales tienen una influencia negativa en los aprendizajes (e.g. Ayala y Sánchez, 2017), especialmente en los estudiantes de bajos ingresos (Marotta, 2019).

- Paragraph's conclusion: multiple evidence on positive impacts
- Paragraph's main point: What do we need to know about them
- Paragraph's idea 1: depends on every context, mention rubric of knowledge

Hincapié et al (2020) - Profesores a prueba nuestros gobiernos asuman compromisos serios para desarrollar e implementar una política sistémica que permita mejorar el prestigio de la profesión y la calidad de los docentes. Esta política debe incluir aspectos a lo largo de la carrera docente, como los siguientes: i) la atracción de mejores candidatos a carreras de Educación y Pedagogía; ii) la mejora de la calidad y pertinencia de la formación inicial; iii) la selección de candidatos a la profesión docente que esté basada en el mérito; iv) un marco que regule la profesión de manera clara, transparente y meritocrática; v) una política de remuneración competitiva que genere incentivos a la excelencia; y vi) una evaluación de los docentes que asegure el mejoramiento continuo (Elacqua et al., 2018; Bruns y Luque, 2015; García et al., 2014).

Al momento de su diseño e implementación, las autoridades educativas se enfrentan a

una multiplicidad de posibilidades, que incluyen la seleccin de instrumentos, la decisin de cmo deben o pueden implementarse y el peso que se le asignar a cada uno dentro del puntaje final, entre otros elementos. Igualmente, una vez que los procesos finalizan, debe decidirse cmo entregar la informacin recolectada a los docentes y, lo que es ms importante an, cmo utilizarla para asegurar la mejora de la labor pedaggica.

Vale la pena aclarar que un paso previo al diseo del sistema de evaluacin y a la eleccin de los instrumentos a ser utilizados tiene que ver con la definicin de lo que debe saber y debe saber hacer un buen docente en cada pas.

La primera pregunta que surge al analizar el tema es cmo podra la evaluacin docente atraer, formar y retener en las aulas a los mejores docentes, y a travs de esto mejorar el aprendizaje de los estudiantes. El grfico 2.1 presenta la teora de cambio, resumiendo los mecanismos a travs de los cuales esto puede llegar a suceder. 1

1 Para esta teora de cambio se asume que los docentes con mayores habilidades pedaggicas y disciplinares efectivamente las despliegan en las aulas y de esta forma pueden mejorar el aprendizaje de sus estudiantes. (esto se puede probar)

Para alcanzar este objetivo es necesario que todos los actores del sistema educativo tengan claro qu es lo que un docente debe saber y debe saber hacer.

- Paragraph's idea 2: disciplinar and pedagogical knowledge are an important part

Hincapié et al (2020) - Profesores a prueba es claro que se espera que los docentes cuenten con conocimientos disciplinares y pedaggicos adecuados a su materia y grado de enseanza, sobre sus estudiantes y su proceso de aprendizaje, sobre su contexto y sobre el uso en el aula de materiales pedaggicos y recursos tecnolgicos, de tal forma que aseguren una enseanza efectiva.

Araujo et al. (2016) para Ecuador, sugieren que alumnos asignados a profesores de alto desempeo dominan el contenido enseado cerca de un 50% ms que uno asignado a un docente de bajo desempeo

- Paragraph's idea 3: how well the teacher translate that into the student - Paragraph's idea 3: socio-emotional and efficiency abilities that impact beyond the class - Paragraph's conclusion: it is useful to know their disciplinar and pedagogical knowledge

Hincapié et al (2020) - Profesores a prueba Todas estas polticas deben tener como objetivo desarrollar y potenciar los conocimientos disciplinares y las habilidades pedaggicas de los docentes.

Las evaluaciones docentes pueden ayudar a identificar las diferencias de desempeo entre los profesores. Adems, el uso adecuado de sus resultados puede otorgar la informacin necesaria para aprovechar al mximo sus fortalezas, buscar superar las falencias, y potenciar la excelencia en la profesin.

- Paragraph's main point: How to identify good teachers

- Paragraph's idea 1: multiple instruments are the key

Bertoni et al (2020) - Concursos Docentes en Latinoamrica Entonces, dada la complejidad del trabajo docente, es recomendable utilizar una multiplicidad de instrumentos para identificar y seleccionar a los mejores docentes.

capacidad de impactar positivamente en sus estudiantes, la evaluacin docente debe basarse en instrumentos que otorguen informacin vlida 2 y confiable 3 cualidades son fundamentales para que la autoridad educativa pueda entender culas son las fortalezas y las debilidades de sus docentes, y pueda llevar adelante las acciones necesarias para potenciarlas o superarlas, respectivamente

2 Cuando los resultados que surgen de su implementacin permiten identificar, mediante

investigaciones rigurosas, docentes altamente efectivos o que tienen un impacto en el aprendizaje de sus estudiantes 3 Cuando los resultados que se obtienen de cada docente reflejan su desempeño típico en clase y no dependen del día en particular en que la información fue recolectada o de la persona que estuvo a cargo de esa recolección.

Hincapié et al (2020) - Profesores a prueba Un instrumento (o un conjunto de instrumentos) de evaluación docente otorga información válida si los resultados que surgen de su implementación permiten identificar, mediante investigaciones rigurosas, docentes altamente efectivos o que tienen un impacto en el aprendizaje de sus estudiantes. Un instrumento (o un conjunto de instrumentos) otorga información confiable cuando los resultados que se obtienen de cada docente reflejan su desempeño típico en clase y no dependen del día en particular en que la información fue recolectada o de la persona que estuvo a cargo de esa recolección. Estas dos cualidades son fundamentales para que la autoridad educativa pueda entender cuáles son las fortalezas y las debilidades de sus docentes y pueda llevar adelante las acciones necesarias para potenciarlas o superarlas, respectivamente.

Si el diseño y la implementación de las evaluaciones docentes no aseguran que los instrumentos y la información que surja de ellas sean válidos y confiables, es poco probable que los recursos invertidos en ellas tengan frutos verdaderos. Incluso si esta primera condición se cumple es importante también el uso que se les da a los resultados que arrojen las evaluaciones.

- Paragraph's idea 2: standardize MCQ evaluations are good enough (at least for the purpose)

Bertoni et al (2020) - Concursos Docentes en Latinoamérica Existe amplia evidencia de que las pruebas de conocimiento y las observaciones estandarizadas de aula son instrumentos relacionados con una mayor efectividad docente (Bruno y Strunk, 2019; Kane et al., 2011; Kane y Staiger, 2012), así como también las entrevistas por parte del director u otro funcionario (Harris y Sass, 2014; Jacob y Lefgren, 2008).

Estudios encuentran que los puntajes obtenidos por los docentes en pruebas de conocimientos están asociados a mayores aprendizajes de los estudiantes (Bietenbeck et al., 2017; Clotfelter et al., 2006; Clotfelter et al., 2007; Hanushek et al., 2017; Metzler y Woessmann, 2012), particularmente en la materia que imparten.

Hincapié et al (2020) - Profesores a prueba Pruebas estandarizadas a los docentes Probablemente, la mayor ventaja de este instrumento es que, una vez su diseño asegure que las preguntas incluidas efectivamente evalúan si el docente cumple o no con los estándares de desempeño requeridos, su implementación es mucho más sencilla y menos costosa que la de los otros instrumentos arriba referenciados. Además, es el único instrumento para el cual hay evidencia causal positiva de la región que muestra cómo su implementación puede efectivamente mejorar el desempeño estudiantil (Ome, 2012; Brutti y Sánchez, 2017; Estrada, 2019).

- Paragraph's idea 3: For now the evidence has been based on proxies, but more can be done

Bertoni et al (2020) - Escasez de docentes en Latinoamérica De acuerdo con Sutch et al., (2016), los principales indicadores que han sido utilizados en la literatura son: 1) porcentaje de vacantes con dificultades para ser llenadas en los concursos docentes; 2) tamaños de clase, ya que la mayoría de los sistemas tienen límites máximos del número de estudiantes por profesor; 3) Porcentaje de docentes sin la preparación necesaria, de acuerdo a los estándares de formación inicial establecidos en la legislación; 4) Porcentaje de docentes sin experiencia; 5) Porcentaje de docentes con contratos temporales, sustitutos o

con certificaciones ad-hoc para ejercer la docencia; 6) Profesores que enseñan una materia o asignatura 1 distinta a aquella en la que se especializaron o en la que obtuvieron su licencia (out of field teachers); 7) Docentes representativos de minoras étnicas, indígenas o de necesidades especiales. Cada uno de estos indicadores tiene ventajas y desventajas para aproximarse al concepto de escasez de docentes.

Bertoni et al (2020) - Escasez de docentes en Latinoamérica En segundo lugar, para Brasil, Chile y Ecuador, se consideran indicadores de idoneidad de los profesores. Según Medeiros et al., (2018), la falta de idoneidad (out-of-field-teaching, en inglés), ocurre cuando el docente enseña una materia en la que no tiene la especialidad y/o el certificado correspondiente

Con respecto a los requisitos generales, para ser docente en Perú es necesario poseer el título de profesor o licenciado en educación, otorgado por una institución de formación docente acreditada en el país o en el exterior (en este último caso, el título debe ser revalidado en el Perú) 16 Además de los requisitos generales también se deben cumplir requisitos específicos, por ejemplo: a) manejar fluidamente la lengua materna de los estudiantes y conocer la cultura local para postular a vacantes de instituciones educativas pertenecientes a educación intercultural bilingüe (EIB); b) acreditar la especialización en la modalidad para postular a vacantes de instituciones educativas pertenecientes a educación básica especial (EBE); y c) se permite enseñar inicialmente a los docentes con título de profesor o de licenciado en educación en la modalidad de educación básica regular (EBR) en el nivel primaria, con estudios concluidos de segunda especialidad en educación inicial y con experiencia mínima de dos (02) años lectivos en el nivel inicial.

Sin embargo se puede tener una medición de idoneidad más precisa, si se conocen los instrumentos. Solo usan el puntaje para medir un proxy de idoneidad

Hincapié et al (2020) - Profesores a prueba últimas décadas del siglo XX. Además, las habilidades pedagógicas y los conocimientos disciplinares de los docentes en la región están por debajo de lo que sugieren los estándares internacionales. Por ejemplo, los docentes de la región dedican 20% menos del tiempo efectivo en clase de lo que estos recomiendan. Es decir, en América Latina, las diferencias en tiempos de enseñanza efectiva de los docentes implican que los estudiantes de la región reciben en promedio un día menos de clase a la semana (Bruns y Luque, 2015). De manera similar, en lo que respecta a los temas disciplinares, pruebas en Perú, Chile y México y estudios internacionales aplicados a los propios docentes indican que sus conocimientos de matemáticas son insatisfactorios (Elacqua et al., 2018).

(solo usan proxies, ver bien el estudio de Elacqua et al., 2018)

- Paragraph's idea 4: Selection has been based on scores

En el contexto de un proceso de evaluación, muchas veces el individuo se ve enfrentado a una prueba “estandarizada”; es decir, una evaluación diseñada de tal manera que las preguntas, las condiciones para ser administrada, los procedimientos de calificación e interpretaciones son consistentes con una manera predeterminada o tipificada. En este contexto, el individuo es expuesto a un instrumento de evaluación cuyos ítems usualmente cumplen con las siguientes tres características: (i) preguntas de opción múltiple o políticas, (ii) preguntas que exhiben categorías nominales, sin un ordenamiento específico y (iii) una respuesta “correcta”; tal y como se observa en la **Figura ??**

(measurement error problem)

- Paragraph's conclusion: standardized evaluation helps to know the knowledge of the teacher

Hincapié et al (2020) - Profesores a prueba Las falencias en la calidad de la formación inicial y las características de los interesados en ingresar a la profesión docente en la región (reseadas por Elacqua et al., 2018), sumadas a la dificultad que tienen los ministerios de educación para remover del cargo a docentes con bajos niveles de desempeño, convierten a las evaluaciones de ingreso a la carrera en un elemento esencial para identificar mejor las características y capacidades de los futuros docentes. En ese sentido, continuar con los procesos de mejora y de implementación adecuada de esta evaluación podrá traer beneficios en la calidad de la fuerza docente en la región. Aunque la evidencia aún es escasa, estudios para Colombia (Ome, 2012; Brutti y Sánchez, 2017) y México (Estrada, 2019) sugieren que estos sistemas de selección (con evaluaciones para ingresar a la carrera) están teniendo algunos impactos positivos en la calidad educativa de los países que los implementan.

- Paragraph's main point: How the results are used

Hincapié et al (2020) - Profesores a prueba La segunda característica necesaria dentro de la teoría de cambio (reflejada en el segundo punto del primer círculo de la cadena) es el uso que se les da a los resultados de la evaluación.

- Paragraph's idea 1: how they are used

Hincapié et al (2020) - Profesores a prueba la Organización para la Cooperación y el Desarrollo Económicos (OCDE) este uso puede tener dos objetivos: i) mejorar las prácticas y las habilidades pedagógicas y/o disciplinares a partir del diagnóstico y la vinculación a programas de desarrollo profesional diseñados para superar los resultados; ii) mejorar la composición y la motivación de la fuerza docente por medio del otorgamiento de bonificaciones, reconocimientos especiales o ascensos para aquellos docentes con resultados excelentes o excluir al docente del sistema o al menos retirarlo de las aulas en los casos en que muestre de manera consistente que no cumple con las condiciones requeridas por la profesión (OCDE, 2009 y 2013). El problema que surge en relación con estos dos objetivos (plasmados en el segundo eslabón del gráfico 2.1) es que son difíciles de lograr con una única herramienta de evaluación. Para poder detectar los aspectos por mejorar de las prácticas y el conocimiento pedagógico y disciplinar, y reconocer la excelencia docente, es necesario que los docentes estén completamente abiertos a revelar sus prácticas y logros y dispuestos a compartirlos con las autoridades

- Paragraph's idea 2: evidence about impacts on the use of the results - Paragraph's conclusion: results can have serious impacts into multiple facets

- Closing thoughts

Bertoni et al (2020) - Concursos Docentes en Latinoamérica en este documento tratamos de responder a dos preguntas fundamentales: ¿cómo identificamos y seleccionamos a los mejores docentes? y ¿cómo los asignamos a las escuelas de una manera eficiente y equitativa?

## 2 Methods

- Paragraph's main point: what method are you using?

- Paragraph's idea 1: IRT and the focus on items

De esta forma, mientras que los modelos para respuestas dicotómicas, tales como Rasch (?), de uno, dos, tres parámetros (?) y cuatro parámetros (?), expresan la probabilidad de elegir la alternativa correcta en función de la "habilidad" del individuo; el **Modelo de Respuesta Nominal (NRM)** y todas sus extensiones (? y ?, capítulo 2), expresa la probabilidad de elegir cada alternativa de la pregunta en función de la misma "habilidad".

A diferencia de los modelos de respuesta graduada (?? y ?, cap?tulo 5), el NRM no se sustenta sobre el concepto de la dicotomizaci?n de las alternativas, que derivan en los umbrales por categor?as caracteristicos de los modelos mencionados; por el contrario, la probabilidad correspondiente a cada alternativa es modelada directamente, implementando una generalizaci?n multivariada del modelo de rasgos latentes log?stico (?, ?).

- Paragraph's idea 2: SEM and the focus on abilities
- Paragraph's idea 3: IRT and SEM equivalence (evidence)

Brown (2015) The potential consequences of treating categorical variables as continuous variables in CFA are manifold: (1) They produce attenuated estimates of the relationships (correlations) among indicators, especially when there are floor or ceiling effects; (2) they lead to pseudofactors that are artifacts of item difficulty or extremeness; and (3) they produce incorrect test statistics and standard errors. ML can also produce incorrect parameter estimates, such as in cases where marked floor or ceiling effects exist in purportedly interval-level measurement scales (i.e., because the assumption of linear relationships does not hold).

Rasch Model with SEM

1. Requires to set the loadings = 1 in all items (there are no evidence that different items should load differently in all sub-factors, if that happen then we can say that an item does not behave good)
2. Thresholds can be transformed into difficulty parameters. They will be from the normal ogive model.

Evidence: It is well known that factor analysis with binary outcomes is equivalent to a two-parameter normal ogive IRT model (e.g., Ferrando Lorenza-Sevo, 2005; Glckner-Rist Hoijtink, 2003; Kamata Bauer, 2008; MacIntosh Hashim, 2003; Moustaki, Jreskog, Mavridis, 2004; Muthn, Kao, Burstein, 1991; Reise, Widaman, Pugh, 1993; Takane de Leeuw, 1987).

Item difficulties have been alternatively referred to in the IRT literature as item threshold or item location parameters. In fact, item difficulty parameters are analogous to item thresholds ( $t$ ) in CFA with categorical outcomes (cf. Muthn et al., 1991).

Item discrimination parameters are analogous to factor loadings in CFA and EFA because they represent the relationship between the latent trait and the item responses.

Muthn (1988; Muthn et al., 1991) has shown that MIMIC models (see Chapter 7) with categorical indicators are equivalent to DIF analysis in the IRT framework (see also Meade Lautenschlager, 2004).

Muthn (1988; Muthn et al., 1991) notes that the MIMIC framework offers several potential advantages over IRT. These include the ability to (1) use either continuous covariates (e.g., age) or categorical background variables (e.g., gender); (2) model a direct effect of the covariate on the latent variable (in addition to direct effects of the covariate on test items); (3) readily evaluate multidimensional models (i.e., measurement models with more than one factor); and (4) incorporate an error theory (e.g., measurement error covariances). Indeed, a general advantage of the covariance structure analysis approach is that the IRT model can be embedded in a larger structural equation model (e.g., Lu, Thomas, Zumbo, 2005).

De esta forma, en el contexto de una evaluaci?n estandarizada, suponemos que  $n$  sujetos responden  $p$  ?tems de opci?n m?ltiple eligiendo **una sola** alternativa de  $m_j$  disponibles, las mismas que pueden variar de ?tem a ?tem y poseen un orden arbitrario. Entonces, el NRM define ***Funciones de Respuestas de las Categor?as del ?tem*** (ICRF, acorde

con  $\theta_i$ ) o Curvas Características de la Alternativas del ítem (IOCC, acorde con  $\theta_i$ ) de la siguiente manera:

$$P_{jk}(\theta_i) = \frac{e^{z_{jk}(\theta_i)}}{\sum_{h=1}^m e^{z_{jh}(\theta_i)}} \quad (1)$$

Donde:

$$z_{jk} = a_{jk}\theta_i + c_{jk} \quad \forall \quad i = 1, \dots, n; \quad j = 1, \dots, p; \quad k = 1, \dots, m_j$$

El parámetro  $\theta_i$  representa la “habilidad” del individuo  $i$ ,  $a_{jk}$  corresponde al parámetro de discriminación de la alternativa  $k$  del ítem  $j$  y  $c_{jk}$  es proporcional a la “popularidad” de la alternativa  $k$  del ítem  $j$ . El vector compuesto por los vectores  $z_{j1}, z_{j2}, \dots, z_{jm_j}$  es usualmente definido como el vector *logit multinomial*. La presente parametrización del modelo es expresada en términos del intercepto y la pendiente de las ICRFs; sin embargo, la literatura utiliza una parametrización que hace la estimación computacionalmente más eficiente.

- Paragraph’s idea 4: what can be gain from this merge

De la parametrización anterior se espera que, al igual que los modelos para respuestas dicotómicas, la ICRF de la alternativa “correcta” sea monotónicamente creciente respecto a la “habilidad”, mientras que la forma de las ICRFs de los distractores dependerá de como la alternativa sea percibida por el evaluado ( $\theta_i$ ). De este modo, se plantea estudiar la formulación, supuestos, características y propiedades del NRM.

De manera complementaria al estudio del modelo, el presente proyecto plantea la estimación de los parámetros de interés a través de simulaciones de **Cadenas de Markov de Montecarlo (MCMC)**, perteneciente a los métodos de inferencia bayesiana. Se elige los métodos bayesianos debido a que: (i) elimina los problemas de no convergencia y estimación impropia de los parámetros encontrados en los procedimientos de máxima verosimilitud conjunta y/o marginal ( $\theta_i$ ), (ii) bajo escenarios en los que la complejidad del modelo incrementa, el método se vuelve más atractivo, pues usa simulaciones en vez de métodos numéricos; (iii) los modelos MCMC se vuelven particularmente útiles cuando los datos son dispersos o cuando es poco probable que la teoría asintótica se mantenga ( $\theta_i$ ); (iv) la flexibilidad y escalabilidad de las soluciones implementadas y (v) una mayor capacidad de recuperación de parámetros de interés, de los cuales existen muchos ejemplos ( $\theta_i$ ,  $\theta_j$ , entre otros).

- Paragraph’s idea 5: What are the difficulties

- Paragraph’s conclusion: SEM/IRT merge provides multiple benefits

- Paragraph’s main point: What data do we have?

Finalmente, el modelo investigado será aplicado a un conjunto de datos reales pertenecientes al sector educativo.

- Paragraph’s idea 1: Standardized MCQ in Peru for multiple purposes

En el actual escenario de la revalorización de la carrera magisterial<sup>1234</sup>, el Ministerio de Educación del Perú (MINEDU) aprobó en el año 2012 e inició la implementación en el año 2014 las evaluaciones a docentes con el propósito de: (i) evaluar las capacidades

<sup>1</sup>Ley N° 28044, Ley General de Educación

<sup>2</sup>Ley N° 29944, Ley de Reforma Magisterial

<sup>3</sup>Decreto Supremo N° 011-2012-ED, que aprueba el Reglamento de La Ley de Educación

<sup>4</sup>Decreto Supremo N° 004-2013-ED, que aprueba el Reglamento de la Ley de Reforma Magisterial, y sus modificaciones



y/o competencias de los docentes nombrados en las especialidades que corresponden a su ense?anza y (ii) revalorizar las escalas salariales de los docentes nombrados. En este contexto, en el a?o 2015, el ministerio aplic? la evaluaci?n de “Ingreso a la Carrera Publica Magisterial y Contratacion Docente” (en adelante **Nombramiento 2015**), la cual permiti? el ingreso de nuevos docentes a la primera de las siete escalas de la carrera magisterial.

- Paragraph’s idea 2: Definition of the sample and variables

El presente proyecto opt? por implementar el modelo investigado en 40 de los 90 ?tems disponibles de Nombramiento 2015, aplicados a 11826 docentes de la especialidad de Matem?tica de la Modalidad de Educaci?n B?sica Regular Nivel Secundaria. El instrumento se encuentra dise?ado para medir un *trazo latente unidimensional* que corresponde a las *competencias pedag?gicas y de especialidad* que los docentes poseen. La elecci?n del modelo se sustent? en que este no solo provee informaci?n acerca de la alternativa elegida (presuntamente “correcta”), sino tambien, permite conocer la “popularidad” con la que el individuo percibe el resto de categor?as disponibles, informaci?n especialmente valiosa para el an?lisis de distractores y validez te?rica de constructo de los ?tems utilizados en el instrumentos de evaluaci?n.

- Paragraph’s idea 3: Composition of the exam - Paragraph’s idea 4: Selection of factors and why - Paragraph’s conclusion: The process can be performed in this data

En conclusi?n, el presente proyecto de tesis estudiar? los supuestos, propiedades y caracter?sticas del Modelo de Respuesta Nominal (NRM) e implementar? la estimaci?n de sus par?metros desde el enfoque de la inferencia bayesiana. Entre los t?picos que adicionalmente ser?n presentados se encuentran: (i) estudios de simulaci?n que comparan la recuperaci?n de par?metros de inter?s entre el m?todo cl?sico de estimaci?n y el bayesiano y (ii) la aplicaci?n a un conjuntos de datos reales del sector educativo, acorde con lo detallado en parrafos previos.

### 3 Thesis objectives

El objetivo general de la tesis consiste en estudiar la formulaci?n, supuestos, caracter?sticas y propiedades del **Modelo de Respuesta Nominal (NRM)** en el contexto de la Teor?a de Respuesta al ?tem (IRT). Del mismo modo, se pretende realizar un estudio de simulaci?n que compare el m?todo cl?sico de estimaci?n del NRM frente a los m?todos bayesianos. Finalmente, se aplicar? el modelo descrito a un conjuntos de datos reales del sector educativo, desde el enfoque de la inferencia bayesiana. De manera espec?fica:

- Se realizar? una extensiva revisi?n de la literatura acerca del modelo de inter?s.
- Se estudiar?n los supuestos, caracter?sticas y propiedades del modelo, desde la perspectiva cl?sica y bayesiana.
- Se implementar?n m?todos de inferencia bayesiana para la estimaci?n de los par?metros de inter?s.
- Se realizar?n estudios de simulaci?n para comprobar la capacidad de recuperaci?n de los par?metros de inter?s por parte del m?todo cl?sico y bayesiano.
- Se aplicar? el modelo de inter?s a un conjunto de datos reales pertenecientes al sector educativo.

## 4 Work plan

Para alcanzar los objetivos del proyecto se consideran los siguientes procedimientos:

1. Revisi3n de literatura acerca de:
  - (a) El Modelo de Respuesta Nominal en la literatura IRT
  - (b) Estimaci3n de los modelos IRT, desde el punto de vista bayesiano
2. Estudio de los supuestos, caracter3sticas y propiedades del modelo, tanto desde el punto de vista cl3sico como bayesiano.
3. Implementaci3n de un programa en aplicativos inform3ticos basados en software libre (R y OpenBUGS).
4. Realizar un estudio de simulaci3n para determinar la capacidad de recuperaci3n de los par3metros de inter3s bajo la inferencia bayesiana.
5. Aplicaci3n del modelo a conjuntos de datos reales.
6. Redacci3n de la tesis.

En el **Cuadro ??**, presentado a continuaci3n, se plantea un cronograma tentativo de actividades:

	2020		2021	
Task	August	December	March	June
Revisi3n de literatura	•	•	•	•
Estudio del modelo	•	•	•	•
Implementaci3n bayesiana	•	•	•	•
Estudio de simulaci3n	•	•	•	•
Aplicaci3n	•	•	•	•
Redacci3n de la tesis	•	•	•	•

Table 1: Cronograma de actividades

**AFDELING**

Straat nr bus 0000  
3000 LEUVEN, BELGIË  
tel. + 32 16 00 00 00  
fax + 32 16 00 00 00  
[www.kuleuven.be](http://www.kuleuven.be)

