



Els efectes dels sistemes de tutoria intel·ligent sobre els aprenentatges: una revisió sistemàtica

**Informe encarregat per l'eLearn Center
de la Universitat Oberta de Catalunya**



Març de 2018

Equip de Treball: Marçal Farré
Miquel Àngel Alegre

Índex

Introducció	1
Què són els Tutors Intel·ligents?	1
Preguntes de recerca	4
Mètode	5
Criteris d'inclusió i exclusió:.....	5
Estratègia de cerca.....	6
Codificació.....	6
Cerca i selecció de documents	7
Síntesis de les revisions i meta-anàlisis identificades	8
Els efectes dels Tutors Intel·ligents segons les meta-anàlisis incloses	8
Variables moderadores dels resultats.....	10
Aplicació del Tutor Intel·ligent	10
Característiques del tutor intel·ligent.....	12
Context en el qual s'implementa el Tutor Intel·ligent	13
Forma com s'avaluen els efectes.....	14
Conclusions de la comparació entre meta-anàlisis	15
Síntesis dels estudis primaris identificats	16
Característiques de la intervenció.....	16
Característiques metodològiques dels estudis.....	17
Anàlisi dels estudis sobre els efectes dels Tutors Intel·ligents.....	18
Anàlisi dels estudis que comparen variacions	21
Conclusions	26
Annex 1: Notes metodològiques	27
Bibliografia	29

Figures i taules

FIGURA 1: ELS TUTORS INTEL·LIGENTS.....	1
FIGURA 2: PROCÉS DE SELECCIÓ DE DOCUMENTS	7
FIGURA 3: DURADA DE LES INTERVENCIIONS AVALUADES.....	17
FIGURA 4: NOMBRE DE PARTICIPANTS EN LES INTERVENCIIONS AVALUADES	18
FIGURA 5: DISTRIBUCIÓ D'11 MIDES D'EFECTE	19
TAULA 1 TERMES CLAU	6
TAULA 2: RESUM DE L'EFECTIVITAT DELS TUTORS INTEL·LIGENTS	9
TAULA 3: MIDES DELS EFECTES SEGONS GRUP DE COMPARACIÓ (MODEL D'EFECTES ALEATORIS).....	10
TAULA 4: MIDES DELS EFECTES SEGONS GRUP DE COMPARACIÓ (MODEL D'EFECTES ALEATORIS).....	11
TAULA 5: MIDES DELS EFECTES SEGONS L'ÚS DE TUTORS INTEL·LIGENTS (MODEL D'EFECTES ALEATORIS)	11
TAULA 6: RESUM DEL RANG DE TEMPS, NIVELL EDUCATIU I ASSIGNATURES DE CADA REVISIÓ	13

INTRODUCCIÓ

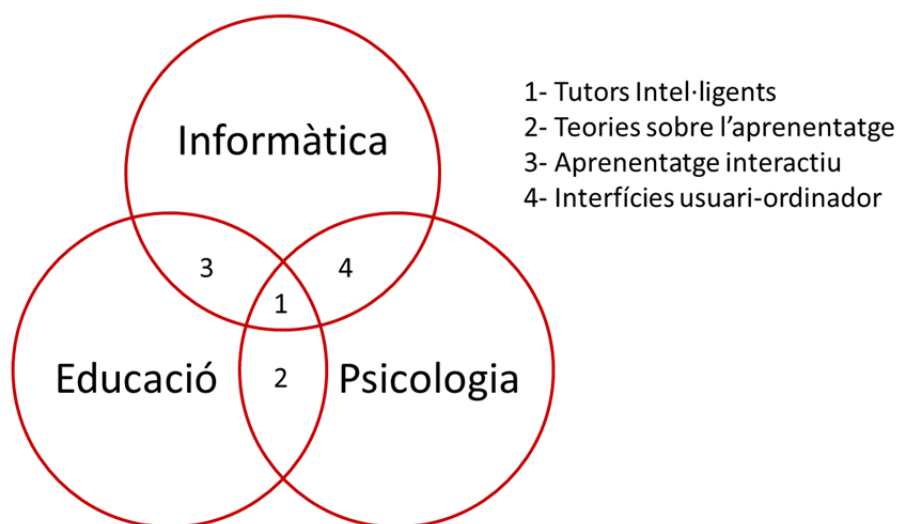
En les darreres dècades l'ús de tecnologies de la informació i la comunicació s'ha expandit en l'àmbit de l'ensenyament. D'entre les seves aplicacions destaquen pràctiques que busquen introduir elements d'automatització en el procés d'ensenyament amb l'objectiu de satisfer i adaptar-se a les necessitats de cada estudiant. Tanmateix, el que sabem sobre l'ús i els efectes d'aquestes tècniques en relació als sistemes tradicionals és encara poc.

El present estudi és una revisió sistemàtica dels efectes d'una tipologia concreta d'intervenció d'automatització de l'ensenyament: els Tutors Intel·ligents.

Què són els Tutors Intel·ligents?

Els Tutors Intel·ligents, *Intelligent Tutoring Systems* (ITS) en anglès, són entorns d'aprenentatge assistit per ordinador que fan servir models computacionals dissenyats per oferir un comportament adaptatiu, que s'ajusta i respon a l'estudiant i a les seves característiques, ritme i necessitats. Per aconseguir tal comportament, els Tutors Intel·ligents incorporen tècniques del camp de la Intel·ligència Artificial, i ho fan tenint en compte teories sobre l'aprenentatge i el comportament humà provinents de diferents camps del coneixement. Així, el disseny dels Tutors Intel·ligents sol ser una tasca interdisciplinària que se situa en la intersecció entre la informàtica, la psicologia i l'educació (Woolf, 2008).

Figura 1: Els Tutors Intel·ligents



Font: Elaborat a partir de Woolf, 2008.

Les teories sobre l'aprenentatge tenen un rol destacat en aquest esquema: ens diuen com s'adquireix, es recorda i es fa servir el coneixement, i representen la base teòrica sobre la qual es construeixen els Tutors Intel·ligents. Per la seva banda, les ciències informàtiques, i més concretament la Intel·ligència Artificial, aporten les eines per

dissenyar sistemes que es comporten de forma intel·ligent, agafant com a model referència el comportament que tindria un tutor humà. Així, per entendre els components d'un Tutor Intel·ligent, és útil mirar d'entendre com funciona el tutoratge humà: Quan els humans ensenyen ho fan disposant d'un volum important de coneixement expert, desplegant diferents estratègies d'ensenyament i tenint en compte les habilitats i estils d'aprenentatge més apropiats per l'estudiant. De la mateixa manera, els Tutors Intel·ligents necessiten disposar d'aquests elements. A continuació s'exposen els components dels Tutors Intel·ligents d'acord amb les definicions de Woolf (2008):

- **Domini de l'àmbit de coneixement:** Representa el coneixement expert de l'àmbit en el que el Tutor ensenya (p. ex. Àlgebra, química, història...).
- **Coneixement de l'estudiant:** Representa el coneixement sobre l'estudiant, el seu comportament i les seves necessitats. Aquest servirà per adaptar el comportament del tutor a les necessitats de l'estudiant.
- **Coneixement de tutoratge:** Representa les estratègies d'ensenyament disponibles, com podrien ser l'ús d'exemples, pistes, analogies, simulacions, animacions, etc.
- **Coneixement de comunicació:** Representa els mètodes per comunicar-se amb l'estudiant, com serien la interfície gràfica, la representació d'un agent o mecanismes de diàleg amb l'estudiant.

Malgrat que a la pràctica no tots els Tutors Intel·ligents disposen dels quatre components descrits, un exemple típic de funcionament que involucri els components mencionats s'iniciaria amb la selecció i presentació d'un problema present en el domini de l'àmbit de coneixement, l'enregistrament i anàlisi de la resposta de l'estudiant, la posterior selecció d'una estratègia de tutoratge que doni resposta a les necessitats de l'estudiant, i la implementació d'aquesta estratègia d'acord a una forma de comunicació adequada. Es tracta d'un cicle en el qual el sistema modela la informació disponible en cada moment per prendre millors decisions.

L'antecedent d'aquesta forma de funcionament el podem trobar en tutors de menor complexitat, com seria el cas dels CAI (*Computer assisted instruction tutors*) o d'altres formes d'*e-learning*. La diferència entre els Tutors Intel·ligents i aquests sistemes és que els primers permeten un nombre infinit de possibles interaccions entre el sistema i l'aprenent (Graesser et al., 2011) mentre que els segons ofereixen un nombre d'interaccions força limitat. Malgrat això, a la pràctica pot ser difícil diferenciar-los. Per això VanLehn (2006) fa una descripció del comportament dels Tutors Intel·ligents en el qual distingeix dues funcions:

- **Outer loop:** selecciona la seqüència de tasques i dona suport en la finalització de cada problema, incloent *feedback* sobre la solució d'acord amb la valoració que el sistema fa de les fortaleses i debilitats de l'estudiant en relació als objectius d'aprenentatge.

- **Inner loop:** suggereix passes dins de cada tasca i guia a l'estudiant a través de suggeriments, pistes i altre tipus de suport mentre l'estudiant està treballant en un problema.

Segons la visió de VanLehn, els Tutors Intel·ligents compten amb l'*inner* i l'*outer loop*, mentre que els tutors de tipus CAI només tenen un *outer loop*. D'aquesta manera, els CAI poden proveir feedback a l'estudiant a partir de les seves respostes, però no al llarg del procés de pensament que s'esdevé entre resposta i resposta.

Al mateix temps, dins del camp dels Tutors Intel·ligents podem trobar-hi una gran varietat de tutors. Una manera habitual de classificar-los és segons les tècniques que empren per modelar el comportament de l'estudiant. Ma et al (2014) presenta quatre categories de Tutors Intel·ligents:

- **Expectation and Misconception Tailoring (EMT):** L'estudiant i el sistema dialoguen fent servir llenguatge natural.
 - Exemple: *AutoTutor*
- **Model Tracing:** Es modela l'habilitat o coneixement que l'estudiant ha d'adquirir a través d'unes regles de producció consistents en operacions i condicions sota les quals aquestes són activades. Quan l'estudiant fa servir una regla de producció, un procediment bayesià anomenat *knowledge-tracing* actualitza la probabilitat de que hagi après correctament l'habilitat o coneixement.
 - Exemple: *Cognitive Tutor*
- **Constraint-Based Modeling (CBM):** Representa el domini del coneixement com un conjunt de restriccions lògiques, relacionant cada limitació amb situacions que podrien donar-se en el camí de solucionar un problema. Quan l'estudiant viola una restricció el sistema detecta l'error i proporciona el feedback corresponent.
 - Exemple: *SQL-Tutor*
- **Bayesian Network Modeling:** Les xarxes bayesianes representen el domini de coneixement a través de múltiples variables interconnectades, i es calcula el valor de cada variable segons les seves connexions amb altres variables. Un exemple seria una llista de variables binàries que representen conceptes apresos i errors de concepció de l'estudiant, i una llista de variables que representen *ítems* d'un test. Agafant els resultats dels tests, el sistema calcula la probabilitat de que l'estudiant posseeixi cadascun dels conceptes a ser apresos o errors de concepció.
 - Exemple: *Andes*

PREGUNTES DE RECERCA

Les tecnologies de l'educació se solen avaluar des de dues perspectives molt diferents: la del seu funcionament i la dels seus efectes. Així, cal distingir entre aquella recerca que mira d'avaluar si la tecnologia funciona de forma adequada (p. ex. el sistema és capaç d'identificar les respostes incorrectes i donar un feedback adequat?), d'aquella que tracta d'estimar els seus impactes (p. ex. els estudiants que reben feedback aprenen més?).

La present recerca se situa en el segon grup, i es pregunta quins són els efectes dels Tutors Intel·ligents sobre els aprenentatges dels estudiants en entorns d'educació post-obligatòria, així com quins aspectes són rellevants per explicar aquests efectes. A més, també es vol saber quin tipus d'intervencions amb Tutors Intel·ligents estan sent avaluades i com s'estan realitzant aquestes avaluacions.

Preguntes de recerca:

- Quin tipus de Tutors Intel·ligents estan sent avaluats/estudiats?
- Quin és l'estat de l'evidència i la seva qualitat?
- Quins són els efectes del Tutors Intel·ligents sobre els aprenentatges i/o els resultats acadèmics?
- Quins factors són rellevants per explicar l'impacte dels Tutors Intel·ligents?

Per altra banda, el present treball no es planteja respondre a preguntes relacionades amb la implementació dels Tutors Intel·ligents, tals com quin tipus de tutors estan sent implementats, quines barreres existeixen a la seva implementació, quan costa el seu desenvolupament o en quina mesura seria viable generalitzar el seu ús.

MÈTODE

Tenint en compte la naturalesa causal de les preguntes de recerca, el que es planteja és la comparació entre els aprenentatges dels estudiants que han estat exposats a un Tutor Intel·ligent amb els aprenentatges que aquests mateixos estudiants haguessin tingut en absència del Tutor Intel·ligent. Aquesta diferència només es pot estimar a través de mètodes experimentals o quasi-experimentals, els quals fan servir grups de comparació per inferir què hagués passat en absència de la intervenció avaluada. Tanmateix, l'exercici que es planteja en aquesta recerca no és el de realitzar un nou estudi experimental o quasi experimental, sinó el d'identificar, revisar i sintetitzar els estudis ja disponibles.

Amb l'objectiu de que la revisió i anàlisi de l'evidència disponible siguin sistemàtics, a continuació fem constar els criteris d'inclusió i exclusió emprats, així com l'estratègia de cerca.

Criteris d'inclusió i exclusió:

Els criteris d'inclusió i exclusió van ser definits en un protocol abans d'iniciar la cerca. A continuació els detallem:

- **Disseny de l'estudi:** S'inclouen dissenys experimentals i quasi experimentals amb almenys un grup de tractament i un de comparació o control. També s'inclouen dissenys pre-post. S'inclouen les revisions sistemàtiques i meta-anàlisis existents amb l'objectiu de poder-ne fer una síntesis separada a la dels estudis primaris.
- **Tipus de participants:** Els estudis inclosos han de centrar-se en entorns d'educació post-obligatòria, incloent l'educació secundària no obligatòria, l'ensenyament superior i la formació corporativa.
- **Tipus d'intervencions:** Només s'inclouen els estudis en què la intervenció avaluada sigui un Tutor Intel·ligent. Per jutjar si és el cas, es tindrà en compte si l'autor la classifica com a tal i si la descripció que fa de les seves funcionalitats s'ajusta a l'exposada en apartats anteriors.
- **Tipologies d'impactes (outcomes):** S'inclouen els estudis que mesurin els efectes dels Tutors Intel·ligents sobre els aprenentatges dels estudiants. A més, s'inclouen també aquells estudis que facin referència a la satisfacció d'estudiants i/o professors.
- **Entorn:** S'inclouen tant estudis que es duguin a terme en entorns presencials com d'ensenyament a distància o una combinació d'ambdós.

Estratègia de cerca

L'estratègia de cerca ha estat dissenyada per identificar tota la recerca disponible sobre els efectes dels Tutors Intel·ligents. Per això, s'han fet servir termes de cerca relacionats amb l'aprenentatge a través d'ordinadors i l'automatització de l'aprenentatge.

Taula 1 Termes clau

Població	Intervenció	Tipus d'estudi	Outcomes
higher education univer* college train* learn*	Collaborative Complex Learning person* learn* intelligent tutoring System artificial tutor cognitive tutor computer tutor intelligent learning environment automat* gamification machine learning	evaluation intervention treatment impact trial experiment control group controlled tria quasi-experiment* random* systematic review	edu* outcomes learning ivolvement successful learning effectiv* autonom* achiev* perform* academic learning satisf*

Les cerques s'han realitzat en les següents bases de dades:

- ERIC - Cerca realitzada el 07/08/2017, 308 documents identificats.
- Jstor - Cerca realitzada el 7/9/2017, 99 documents identificats.
- Dialnet - Cerca realitzada el 6/9/2017, 21 documents identificats.
- Redinet - Cerca realitzada el 7/9/2017, 16 documents identificats.

La cerca s'ha restringit a les publicacions posteriors al 2013, per tal de centrar la revisió en l'evidència més recent. El total de documents localitzats és de 444, incloent articles acadèmics i literatura gris.

Codificació

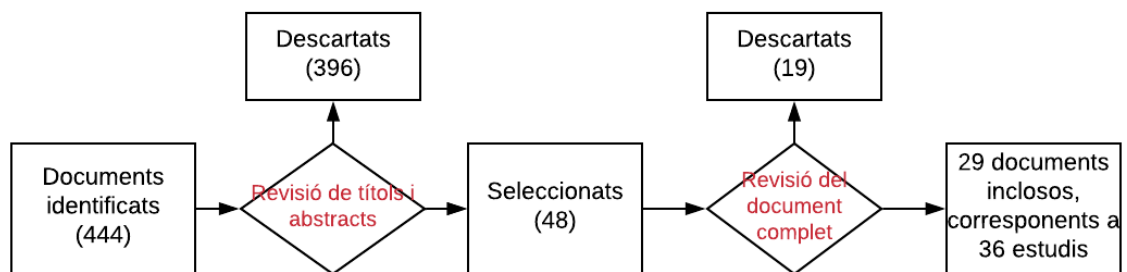
El contingut dels estudis inclosos ha estat codificat per tal d'extreure'n la informació més rellevant. La codificació s'ha realitzat d'acord amb el protocol prèviament dissenyat per cobrir les principals característiques de cada estudi, incloent qüestions com la intervenció avaluada, la metodologia de l'estudi, els participants i els efectes.

CERCA I SELECCIÓ DE DOCUMENTS

Les cerques realitzades han resultat en 444 documents identificats, 396 dels quals han estat descartats després de revisar-ne els títols i resums, i 19 després de revisar el text complet. La major part de les exclusions han estat degudes a que l'estudi no feia referència a intervencions amb Tutors Intel·ligents, però també s'han exclòs estudis pel fet de que no n'avaluaven els efectes o ho feien per nivells d'ensenyament inferiors a l'educació secundària post-obligatòria.

Els documents finalment inclosos són 29. Tenint en compte que alguns documents contenen diversos estudis, la revisió compta amb 36 estudis inclosos.

Figura 2: Procés de selecció de documents



Dels 36 estudis inclosos, 4 són revisions sistemàtiques d'evidència i 32 són estudis primaris experimentals, quasi-experimentals o pre-post. En la majoria dels casos es tracta de publicacions en revistes científiques (31), mentre que només una minoria són informes, documents relatius a conferències o altres documents de literatura gris (4), i en un cas és una tesi. Tots els documents estan escrits en anglès.

En el següent apartat es sintetitzen les quatre revisions d'evidència identificades. A més, s'ha decidit incloure una cinquena meta-anàlisi, anterior al període pel qual s'ha realitzat la cerca (posterior a 2013), per tal de guanyar una major perspectiva temporal sobre l'evolució de l'evidència disponible.

SÍNTESIS DE LES REVISIONS I META-ANÀLISIS IDENTIFICADES

Si bé es considera que els primers Tutors Intel·ligents apareixen a la dècada de 1970 (Woolf, 2009), no és ben bé fins la segona dècada del segle XXI que es comencen a trobar revisions sistemàtiques dels seus efectes sobre l'aprenentatge. Al 2011 VanLehn resumia el que fins aleshores eren les creences sobre els efectes dels *Computer Assisted Instruction* (CAI), els Tutors Intel·ligents i el tutoratge humà (VanLehn, 2011). Així, es creia que els CAI milloraven els resultats dels exàmens 0,3 desviacions típiques, els Tutors Intel·ligents ho feien una desviació típica, i els tutors humans dues desviacions típiques. Aquestes creences són atribuïdes a estudis influents citats per l'autor, però en cap cas a revisions sistemàtiques. Entre aquests estudis, destaca l'article de Bloom (1984) qui va plantejar el problema de les dues sigmes.

Problema de les dues sigmes:

Bloom va definir el problema de les dues sigmes com la cerca d'un mètode que igualés la millora dels aprenentatges de dues desviacions típiques, atribuïda a la combinació d'un tutor humà amb la metodologia del *mastery learning*. Bloom basava aquesta atribució en els estudis de dos estudiants seus (Anania, 1981; Burke, 1983) en els que es comparaven els resultats d'un grup control convencional amb dos grups que utilitzaven la metodologia del *mastery learning*, un dels quals, a més, comptava amb l'assistència d'un estudiant universitari format com a tutor. Aquests estudis mostraven com els resultats d'aplicar la metodologia del *mastery learning*, augmentava els resultats 1,2 desviacions típiques, mentre que afegir-hi un tutor feia augmentar els resultats fins a 2 desviacions típiques respecte el grup de control.

El problema de les dues sigmes ha marcat en bona mesura les expectatives en el camp dels Tutors Intel·ligents, els quals han estat dissenyats per mimetitzar el comportament d'un tutor humà expert (Graesser et al, 2011; Woolf, 2009), esperant aconseguir resultats similars als reportats per Bloom i els seus estudiants. Tanmateix, VanLehn (2011) explora els efectes del tutoratge humà a partir de 10 estudis, reportant un efecte del tutoratge humà força menor al de Bloom, concretament de 0,79 desviacions típiques. La revisió de VanLehn, a més, va ser pionera per sintetitzar evidència sobre diferents tipologies d'instrucció: tutoratge humà, dos tipologies de Tutors Intel·ligents (*step-based* i *sub-step based*), altres formes de tutoratge amb suport informàtic (p. ex. CAI o CBI) i instrucció convencional sense tutoratge. Des d'aleshores s'han realitzat noves revisions d'evidència que permeten avui fer un anàlisi més informat sobre els efectes dels Tutors Intel·ligents.

Els efectes dels Tutors Intel·ligents segons les meta-anàlisis incloses

En aquest apartat es sintetitzen narrativament els resultats de les cinc meta-anàlisis incloses en la present revisió. No es calcula l'efecte conjunt de totes elles, però sí que es comparen els efectes reportats per cadascuna. Per això es fa servir el concepte de la mida de l'efecte. En l'annex 1 s'explica aquest concepte així com els models emprats per calcular la mitjana ponderada entre diferents mides d'efecte.

La cerca ha identificat 4 revisions publicades posteriorment a l'any 2013. A més, s'ha decidit incloure la ja mencionada revisió de VanLehn (2011), tenint en compte que és, si no la primera, una de les primeres revisions sistemàtiques de l'evidència sobre els efectes dels tutors Intel·ligents. Cal mencionar que s'ha trobat una segona meta-anàlisi d'Steenbergen-Hu i Cooper, però s'ha exclòs degut a que es centra únicament en els efectes dels Tutors Intel·ligents per a la població d'estudiants en cursos inferiors a la secundària post-obligatòria.

D'entre les cinc revisions incloses, quatre reporten mides de l'efecte positives i estadísticament significatius, que se situen entre 0,35 i 0,58, i una, la revisió de la *What Works Clearinghouse*, no calcula la mida de l'efecte però reporta efectes mixtos: d'entre els sis estudis que inclou, un mostra efectes negatius, un altre mostra efectes positius i quatre estudis mostren efectes indeterminats.

La següent taula resumeix aquests resultats indicant el nivell educatiu en el qual s'han realitzat els estudis inclosos en cada revisió, així com el nombre de grups de comparació que han servit per calcular la mida de l'efecte mitjà.

Taula 2: Resum de l'efectivitat dels Tutors Intel·ligents

Meta-anàlisi	Nivell educatiu	N de comparacions	Mida de l'efecte ¹	Interpretació de l'efecte
VanLehn (2011)	Sense restriccions	54	d=0,58	Mitjà
What Works Clearinghouse (2013)	<i>High School</i>	6	---	Inconsistent
Steenbergen-Hu i Cooper (2014)	Universitari	37	g=0,35	Petit/mitjà
Ma et al. (2014)	Sense restriccions	107	g=0,41	Petit/mitjà
Kulik & Fletcher (2016)	Sense restriccions	63	g=0,49	Petit/mitjà

A partir d'aquests resultats podem observar que els Tutors Intel·ligents tendeixen a tenir uns efectes moderadament positius sobre els aprenentatges dels estudiants. La revisió de la *What Works Clearinghouse* (2013) és l'única evidència que posa en dubte aquesta afirmació. Es tracta d'un informe que reporta 6 estudis primaris sobre intervencions que comparteixen unes característiques similars: s'han dut a terme en assignatures de matemàtiques, en *High Schools* dels Estats Units d'Amèrica i en totes elles s'ha fet servir un *Cognitive Tutor*.

En els següents apartats fem un anàlisi més aprofundit dels resultats, explorant quines condicions poden estar associades amb una major efectivitat dels Tutors Intel·ligents i quins efectes diferencials es poden produir per diferents perfils d'estudiants.

¹ A l'Annex 1 s'explica el concepte de mida de l'efecte.

Variables moderadores dels resultats

Hi ha moltes explicacions possibles sobre què fa que els Tutors Intel·ligents tinguin uns efectes o uns altres sobre els aprenentatges. Partint de la identificació de possibles variables explicatives, aquestes s'han agrupat en les següents categories:

- **Implementació del Tutor Intel·ligent:** Té a veure en com s'implementa el Tutor Intel·ligent, incloent l'ús que se'n fa (per fer els deures, per substituir o complementar el professor, per preparar els exàmens...), el disseny de la intervenció (la durada, intensitat, preparació dels professors, etc.) i la situació respecte la qual es compara.
- **Característiques del Tutor Intel·ligent:** Inclou qüestions com la tècnica per modelar el comportament de l'estudiant, les estratègies d'ensenyament, si té en compte l'estat emocional de l'estudiant, com presenta la informació, etc.
- **Context en el qual s'implementa:** Inclou qüestions com el país, el nivell educatiu, el perfil dels estudiants, si es tracta d'estudis presencials o a distància o l'assignatura en la qual es decideix implementar el Tutor Intel·ligent
- **Forma com s'avaluen els efectes:** Inclou el disseny de l'estudi, la forma com s'han seleccionat els participants, la mida de la mostra, els instruments emprats per mesurar els aprenentatges dels estudiants, etc.

A continuació explorem aquestes dimensions a través de les meta-anàlisis mencionades.

Implementació del Tutor Intel·ligent

Els resultats de la implementació dels Tutors Intel·ligents varien segons la intervenció amb la que es compara. Steenbergen-Hu i Cooper (2014) en la seva revisió van definir set categories segons la intervenció de comparació, trobant un efecte major respecte l'absència de tractament, l'aprenentatge autònom i la lectura impresa, mentre que els efectes respecte l'aula tradicional o aprenentatge assistit per ordinador són moderats o petits, i són negatius però no estadísticament significatius en relació al tutoratge humà:

Taula 3: Mides dels efectes segons grup de comparació (model d'efectes aleatoris)

Intervenció de comparació	Mida de l'efecte (g)
Absència de tractament	0,9
Aprenentatge autònom	0,82
Lectura impresa	0,50
Aula tradicional	0,38
Aprenentatge assistit per ordinador	0,33
Lectura de materials digitalitzats	0,25 (no significatiu)
Tutoratge humà	-0,25 (no significatiu)

Font: Steenbergen-Hu i Cooper, 2014.

De forma consistent amb aquests resultats, Ma et al. (2014) comparen els efectes segons la mida del grup de comparació i la tipologia d'intervenció, mostrant com els Tutors Intel·ligents són més efectius comparats amb grups grans i ho són menys comparats amb grups petits o amb instrucció individualitzada. A la vegada, mostra efectes significativament positius en relació a altres intervencions que fan servir ordinadors com a suport (*Computer-Based Interventions*) i a l'ús dels llibres de text o exercicis individuals.

Taula 4: Mides dels efectes segons grup de comparació (model d'efectes aleatoris)

Grup de comparació	Mida de l'efecte (g)
Instrucció humana en grups grans	0,44
Instrucció humana en grups petits	0,05 (no significatiu)
Instrucció humana individualitzada	-0,11 (no significatiu)
<i>Computer Based Interventions</i>	0,57
Llibres de text o d'exercicis individuals	0,36

Font: Ma et al., 2014.

Independentment del grup de comparació, també resulta interessant veure els efectes segons quin ús s'ha fet del Tutor Intel·ligent en les intervencions avaluades. Les meta-anàlisis de Steenbergen-Hu i Cooper (2014) i Ma et al (2014) fan aquest exercici a partir d'unes mateixes categories d'intervenció.

Taula 5: Mides dels efectes segons l'ús de Tutors Intel·ligents (model d'efectes aleatoris)

Ús del Tutor Intel·ligent	Steenbergen-Hu & Cooper (2014)		Ma et al. (2014)	
	N d'estudis	Mida de l'efecte (g)	N d'estudis	Mida de l'efecte (g)
Principal mitjà d'instrucció	14	0,30	35	0,37
Integrat en les activitats de classe	4	0,29	15	0,33
Assistir en activitats	2	0,43	24	0,47
Suplementar les instruccions de classe	3	0,73	8	0,43
Suport en els deures	2	0,21	15	0,45
No reportat	---	---	10	0,48

Font: Steenbergen-Hu i Cooper (2014) i Ma et al. (2014).

Segons aquests resultats, sembla que els Tutors Intel·ligents són més efectius quan es fan servir per assistir en activitats de classe o per suplementar les instruccions de classe que no pas com a principal mitjà d'instrucció. Tot i així, en tots dos casos les diferències observades entre categories no són estadísticament significatives.

Per altra banda, més enllà de l'ús del Tutor Intel·ligent i amb quin tipus d'intervenció es compara, hi ha qüestions rellevants com la durada i la intensitat de la intervenció. En aquest sentit, Ma et al. (2014) observen com aquelles intervencions més duradores tendeixen a produir efectes majors. Així, aquelles intervencions de durada igual o

inferior a un mes tenen un efecte (g) de 0,34 sota el model d'efectes aleatoris (*random-effect model*), mentre que aquelles de durada superior a un mes tenen un efecte de 0,38. Tanmateix, els propis autors alerten que aquestes diferències han de ser interpretades amb cautela degut a que un bon nombre d'estudis no reporten la durada de la intervenció.

Finalment, cal tenir en compte que molt pocs estudis mesuren i reporten informació sobre l'adequació de la implementació que avaluen. Kulik i Fletcher (2016) en la seva revisió comparen estudis en els que s'han reportat implementacions fortes i febles. Segons aquests estudis l'efecte medià (ES) en implementacions fortes és de 0,44 i -0,01 en implementacions febles.

Característiques del tutor intel·ligent

Tal com hem vist en apartats anteriors, si bé els Tutors Intel·ligents comparteixen una sèrie de característiques bàsiques entre ells, cadascun compta amb uns atributs únics, com pot ser el model emprat per analitzar i incorporar el coneixement sobre l'estudiant, el repertori d'estratègies d'ensenyament disponibles, el domini de coneixement expert o la forma com es comunica amb l'estudiant. Quatre de les cinc meta-anàlisis incloses en la present revisió han explorat si els efectes del Tutor Intel·ligent varien segons algunes d'aquestes característiques.

Ma et al (2014) comparen diferents tipologies de Tutors Intel·ligents segons el model que fan servir per preveure el comportament de l'estudiant. Així, mostren com els *Constraint-Based modeling* (0,24) i els *Expectation and Misconception Tailoring* (0,34), no produeixen efectes (g) estadísticament significatius, mentre que sí que ho fa el *Model Tracing* (0,35) i el *Bayesian networks* (0,54). Les diferències entre tipologies no són significatives en un model d'efectes aleatoris (*random effect*) però sí que ho són d'acord al model d'efectes fixes (*fixed effects*).

Per la seva banda, Steenbergen-Hu i Cooper (2014) reporten que s'han avaluat una gran varietat de tutors, sumant un total de 22 tipologies diferents. Tot i així, només quatre d'elles sumen més de tres avaluacions, i malgrat que hi ha diferències observades en els seus efectes, aquestes no són estadísticament significatives.

VanLehn (2011) en al seva revisió classifica els Tutors Intel·ligents en dues grans categories segons el nivell de detall o granularitat en la resposta i feedback que dona a l'estudiant:

- **Step based:** aporten pistes i explicacions en cada passa que l'estudiant fa per resoldre un problema.
- **Substep based:** proveeixen aquesta informació a un nivell inferior de granularitat.

Contràriament a les creences més esteses, VanLehn reporta que un menor grau de granularitat produeix millors resultats. Així, els tutors *step-based* tindrien un efecte (d) sobre els aprenentatges de 0,76, mentre que els *substep-based* el tindrien de 0,40. Paradoxalment, aquests resultats suggereixen que els Tutors Intel·ligents més antics i

simples són més efectius, mentre que els tutors més moderns no es mostren més efectius que un tutor de tipologia *Computer Assisted Instruction (CAI)*.

En relació a aquesta darrera qüestió, Kulik i Fletcher (2016) observen que alguns dels estudis que avaluen tutors *substep-based* compten amb grups de control que anomenen “no convencionals”, els quals no fan servir el Tutor Intel·ligent però fan servir materials que se’n deriven. Havent exclòs abans aquests estudis, la diferència entre els efectes (ES) d’ambdues tipologies desapareix (0,60 pels *step-based* i 0,63 pels *substep-based*). Així, Kulik i Fletcher atribueixen els efectes reduïts que VanLehn troba pels tutors *substep-based* al tipus de grups control emprat i no a les característiques del tutor.

Context en el qual s’implementa el Tutor Intel·ligent

L’anàlisi de l’entorn en el qual s’implementa el Tutor Intel·ligent és una qüestió rellevant per entendre sota quines condicions pot ser efectiu. De les cinc revisions sistemàtiques incloses, algunes d’elles han restringit els estudis revisats segons el nivell educatiu o l’assignatura. Així, VanLehn (2011) se centra en assignatures de l’àmbit de les STEM (Ciències, tecnologia, enginyeries i matemàtiques), Steenbergen-Hu i Cooper (2014) només inclouen estudis realitzats en universitats i la revisió de la *What Works Clearinghouse* (2013) se centra en estudis duts a terme en centres de secundària post-obligatòria en l’assignatura de matemàtiques. Ma et al (2014) i Kulik i Fletcher (2016) no han inclòs restriccions pel que fa al nivell educatiu ni a les assignatures.

Taula 6: Resum del rang de temps, nivell educatiu i assignatures de cada revisió

Revisió	Rang de temps	Nivell educatiu	Assignatures
VanLehn (2011)	1975-2010	Sense restriccions	STEM
Steenbergen-Hu i Cooper (2014)*	1990-2011	Universitari	Sense restriccions
Ma et al. (2014)*	Anterior a 2013	Sense restriccions	Sense restriccions
What Works Clearinghouse (2013)	Anterior a 2013	High School	Matemàtiques
Kulik & Fletcher (2016)	Anterior a 2015	Sense restriccions	Sense restriccions

Quan la revisió inclou avaluacions dutes a terme en diferents entorns, és possible analitzar si existeixen diferències en els efectes observats. Steenbergen-Hu i Cooper (2014) reporten efectes majors en el cas de les matemàtiques i estadística (0,59 i 0,46 respectivament), i més reduïts en informàtica (0,31), física (0,27) i gestió d’empreses (0,26). Tot i així, les diferències entre elles no són estadísticament significatives.

Per la seva banda, Ma et al (2014) troben efectes (g) grans i mitjans en el cas de les humanitats (0,63), biologia (0,59) i informàtica (0,51), i reduïts en física (0,38), matemàtiques (0,35), llengües (0,34) i química (0,16). En aquest cas, les diferències

entre les categories tampoc no són significatives sota el model d'efectes aleatoris però sí que ho són sota el model d'efectes fixes.

En relació a aquesta qüestió, Steenbergen-Hu i Cooper (2014) apunten que s'ha produït recerca extensa en àmbits com les matemàtiques, la física o la informàtica, camps en els que les fronteres entre allò correcte i incorrecte són clares, però que encara hi ha poca evidència sobre els efectes dels Tutors Intel·ligents en camps en que aquesta frontera no és tant clara.

Pel que fa al nivell educatiu, Ma et al (2014) reporten diferències molt reduïdes i no significatives entre nivells. Tanmateix, Steenbergen-Hu i Cooper en una revisió anterior centrada exclusivament en estudiants de k-12 (Steenbergen-Hu i Cooper, 2013) reportaven uns efectes (g) en les assignatures de matemàtiques d'entre 0,01 i 0,09, substancialment inferiors als efectes sobre estudiants universitaris en assignatures de matemàtiques (0,35) reportats en la revisió inclosa en el present estudi (Steenbergen-Hu i Cooper, 2014). Finalment, Kulik i Fletcher reporten una correlació positiva entre la mida dels efectes i el nivell educatiu en el qual es desenvolupa la intervenció, sent els efectes (ES) majors en estudis superiors a la secundària (0,75) que en la primària i secundària (0,44).

Finalment, els estudis també es poden dividir segons si s'han realitzat en un entorn creat a propòsit per realitzar l'estudi o si s'ha fet en un entorn educatiu prèviament existent, com podria ser una classe d'institut o universitat. El primer grup són el que es coneix com estudis de laboratori, mentre que el segon són estudis d'aula o de classe. Ma et al. (2014) troben 26 estudis de laboratori i 81 d'aula, i mostren efectes positius en ambdós casos però majors en els estudis d'aula. Les diferències entre els efectes d'ambdues categories són significatives només sota el model d'efectes fixes.

Forma com s'avaluen els efectes

La forma com una intervenció és avaluada pot tenir conseqüències en els seus efectes estimats. És per aquest motiu que les revisions sistemàtiques i les meta-anàlisis prenen una sèrie de precaucions en la selecció d'estudis i el seu anàlisi, tractant d'evitar tot tipus de biaixos i amenaces a la credibilitat dels resultats. Un exercici interessant que realitzen algunes meta-anàlisis és comparar si els resultats difereixen segons les característiques metodològiques dels estudis primaris inclosos en la revisió.

Kulik i Fletcher (2016) en la seva meta-anàlisi mostren com el tipus de test que es fa per mesurar els aprenentatges dels estudiants és una variable correlacionada positivament amb la mida dels efectes del Tutor Intel·ligent. Així, els efectes (ES) tendeixen a ser majors en estudis que empen tests desenvolupats localment (0,73), i menors en avaluacions que fan servir tests estandarditzats (0,13), mentre que avaluacions que fan servir una combinació dels dos tipus de tests obtenen efectes intermitjos (0,45). Segons Kulik i Fletcher (2016), això explicaria el fet que els estudis més antics tendeixin a mostrar efectes més grans, ja que aquests estudis també tendeixen a emprar tests localment desenvolupats mentre que els estudis més

recents, per contra, basen les seves conclusions en tests estandarditzats. Prèviament, Ma et al (2014) havien reportat diferències molt menors entre estudis amb tests desenvolupats localment i estandarditzats: Sota un model d'efectes aleatoris els primers obtenen efectes (g) de 0,41 i els segons de 0,42, sent la diferència no significativa, mentre que sota un model d'efectes fixes els primers obtenen efectes de 0,40 i els segons de 0,27, sent aquest cop la diferència a favor dels tests locals significativa.

Els defensors dels "tests locals" argumenten que aquests són més propicis a estar alineats amb els objectius del programa o intervenció que s'avalua. Per contra, els defensors dels tests estandarditzats defensen que aquests garanteixen absència de biaixos, ja que han estat desenvolupats per terceres parts que no participen en el programa avaluat i per tant no poden preveure si estaran afavorint a un grup o altre grup.

Relacionat amb els instruments de mesura, una altra qüestió rellevant és quin tipus de coneixement esperem observar en l'estudiant. Ma et al (2014) reporten que els estudis que fan servir Tutors Intel·ligents per l'adquisició de coneixement de tipus procedimental i declaratiu tenen efectes superiors que aquells que fan servir els Tutors Intel·ligents amb l'objectiu d'adquirir només coneixement procedimental o declaratiu. El coneixement declaratiu és el coneixement conscient sobre fets, conceptes o idees conegudes, mentre que el procedimental és el coneixement implícit que té a veure en "com fer les coses".

A més, Kulik i Fletcher (2016), reporten una relació forta entre la mida de l'efecte efecte i de la mostra de participants, sent els efectes més reduïts en estudis que compten amb un major nombre de participants.

Finalment, pel que fa al disseny de l'estudi, Ma et al (2014) reporten trobar mides d'efectes més grans en estudis de dissenys quasi-experimentals que no controlen estadísticament les diferències prèvies entre grups. Tanmateix, demanen precaució en la interpretació dels resultats atès a que un volum important d'estudis no reporta explícitament el disseny de l'estudi.

Conclusions de la comparació entre meta-anàlisis

Els Tutors Intel·ligents mostren consistentment resultats positius respecte totes les intervencions alternatives analitzades excepte els tutors humans i els grups reduïts. En relació als primers, VanLehn (2011), Steenbergen-Hu i Cooper (2014) i Ma et al. (2014) coincideixen en trobar diferències estadísticament no significatives, i pel que fa als grups reduïts, Ma et al (2014) mostren diferències reduïdes i estadísticament no significatives a favor dels Tutors Intel·ligents. L'evidència acumulada també ens mostra com els Tutors Intel·ligents són efectius de forma consistent emprats de diverses maneres, ja sigui com a principal mitjà d'instrucció, com una eina integrada en les activitats a l'aula o com a suport per fer els deures.

Com hem vist en els apartats anteriors, hi ha un bon nombre de variables potencialment rellevants per explicar efectes dels Tutors Intel·ligents. Tanmateix, cal

tenir certa precaució en la interpretació de la importància d'aquestes variables, ja que en molts casos es troben correlacionades entre sí i és difícil saber què explica les diferències observades. Per exemple, els tests estandarditzats se solen fer servir més en estudis que tenen una mostra elevada, duts a terme en centres de *middle, junior* i *high school* dels Estats Units, i en els que el tutor implementat sol ser el *Cognitive Tutor*. Kulik i Fletcher (2016) reporten que aquests estudis estimen uns efectes menors que la resta, tanmateix, com ells mateixos alerten, és difícil saber a quines característiques concretes podem atribuir aquests resultats. De la mateixa manera, els estudis més antics reporten mides d'efectes substancialment superiors, però no sabem del cert a què ho podem atribuir.

Així, és important continuar acumulant evidències sobre els efectes dels Tutors Intel·ligents en relació a intervencions alternatives, però també cal tenir en compte estudis que comparin diferents variacions directament entre elles, ja siguin en el disseny del tutor, en l'ús que se'n fa, la durada de la intervenció, la seva intensitat, etc.

Amb l'objectiu d'aprofundir en l'anàlisi de quines són les característiques que fan que un Tutor Intel·ligent sigui efectiu, en el següent apartat fem una síntesis narrativa dels resultats dels estudis primaris identificats en la nostra revisió sistemàtica.

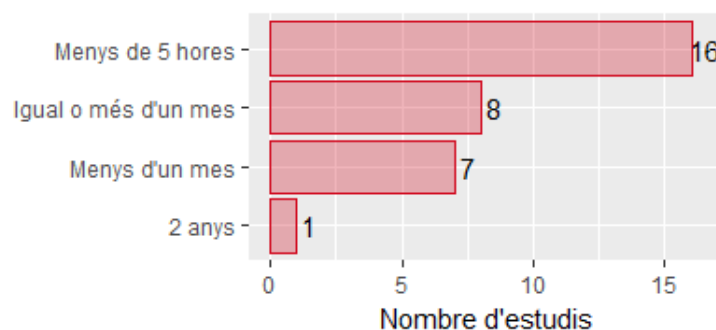
Síntesis dels estudis primaris identificats

Entre els 32 estudis primaris identificats i inclosos en la present revisió, 18 s'han dut a terme a l'Amèrica del Nord (15 dels Estats Units i 3 de Canadà), sis a Europa (un estudi en cadascun dels següents països: Regne Unit, Espanya, Grècia, França, Portugal i Noruega), tres a Oceania (tots a Nova Zelanda, un dels quals comparteix part de la mostra de participants amb Oman), tres a Àsia (un estudi en cadascun dels següents països: Taiwan, Corea, Malàisia), un a Àfrica (Sudàfrica), i en un cas no queda clar on s'ha fet. A continuació es descriuen les característiques de les intervencions avaluades i la metodologia que han fet servir per estimar-ne els efectes.

Característiques de la intervenció

La meitat dels estudis (16) avaluen intervencions de durada molt curta, generalment consten d'entre una i quatre sessions que en total no sumen més de cinc hores. En altres casos la intervenció inclou més sessions i pot arribar a durar varies setmanes sense arribar al mes (8) o superar el mes de durada (7). Només s'ha trobat un cas en què la intervenció avaluada s'estengui més enllà d'un curs acadèmic, es tracta d'un estudi que ha avaluat els efectes d'implementar un Tutor Intel·ligent durant dos anys.

Figura 3: Durada de les intervencions avaluades



Entre els diferents models de Tutor Intel·ligent avaluats destaca el *Cognitive Tutor*, emprat en set dels 32 estudis. Es tracta d'un tutor desenvolupat pel *Pittsburgh Science of Learning Center*, i que fa uns anys es feia servir al voltant de 2000 centres escolars dels Estats Units d'Amèrica (Graesser et al., 2011). De forma menys destacada, la present revisió ha identificat quatre estudis en què es fa servir el Tutor Intel·ligent *Dragoon*, vinculat a l'Arizona State University. En la majoria dels casos, els autors de l'estudi estan d'alguna manera vinculats amb el desenvolupament del Tutor Intel·ligent avaluat (24), ja sigui perquè formen part de l'equip que el va desenvolupar inicialment o perquè n'han fet alguna modificació o adaptació. Tanmateix, sovint l'estudi no ho explicita, per la qual cosa és possible que la proporció sigui major.

En la majoria dels casos el Tutor Intel·ligent s'ha implementat dins de l'aula (26), en front d'una minoria de casos en què s'ha fet servir dins i fora de l'aula (2), com a suport fora de l'aula (2), o no quedava prou clar quin ús se'n va fer (2). Tanmateix, en tots els casos es tracta d'entorns d'ensenyament presencial.

Per altra banda, la informació reportada sobre el contingut de la intervenció avaluada en alguns casos no és gaire detallada, i només s'identifica un estudi que hagi mesurat de forma sistemàtica el grau de fidelitat de la implementació de la intervenció respecte el seu disseny inicial.

Finalment, pel que fa al nivell educatiu, la majoria d'intervencions s'han dut a terme en estudis universitaris (17), seguides de secundària no obligatòria (11) i formació professional (3). No s'han trobat estudis en entorns de formació corporativa. En la gran majoria dels casos es tracta d'intervencions desenvolupades en assignatures de l'àrea de les matemàtiques (11) i la informàtica (11), seguides de biologia (3), llengües (3), física (2), química (1).

Característiques metodològiques dels estudis

Dels 32 estudis inclosos 21 són de tipus experimental, deu quasi-experimental i un és un disseny pre-post. Deixant de banda l'estudi pre-post, en la gran majoria dels casos l'assignació a la intervenció o al grup control es fa a nivell d'estudiant (25), mentre que en alguns casos es fa a nivell de classe (5) i només en un cas a nivell de centre educatiu. En la majoria dels estudis el mètode d'assignació és aleatori (19), mentre que

en la resta es fan servir altres mètodes que tracten d'assegurar la comparabilitat entre grups o no queda clarament explicat. Tanmateix, en la majoria de casos en què l'assignació s'ha fet de forma aleatòria no es reporta informació sobre com s'ha realitzat (per ordinador, sorteig, etc.).

Si ens fixem en el nombre de participants en les intervencions avaluades, els valors es troben entre 19 i 13.445 participants. Tal com mostra el següent diagrama de caixes, la mediana es troba en 52,5 participants i menys d'una quarta part dels estudis superen els 100 participants. Només hi ha dos estudis que compten amb més de 145 participants, un d'ells amb 400 i l'altre 13.445 participants. Aquest darrer no s'ha representat en el gràfic per tal de no distorsionar-ne l'escala.

Figura 4: Nombre de participants en les intervencions avaluades



La majoria dels estudis mesuren els efectes del Tutor Intel·ligent sobre els aprenentatges de l'estudiant durant la pròpia intervenció o de forma molt immediata a la seva finalització, i pràcticament no hi ha estudis que facin seguiment dels aprenentatges dels estudiants al cap d'un temps.

Finalment, entre els 32 estudis primaris inclosos, 16 tracten d'estimar els efectes dels Tutors Intel·ligents mentre que n'hi ha 17 que tracten d'estimar els efectes de diferents variacions d'aquests. Hi ha un estudi que es troba en els dos grups.

A continuació descrivim de forma separada les dues tipologies d'estudis incloses en la present revisió.

Anàlisi dels estudis sobre els efectes dels Tutors Intel·ligents

D'entre els 16 estudis primaris inclosos que estudiaven els efectes dels Tutors Intel·ligents s'han pogut extreure les mides dels efectes d'11 d'ells. La següent figura mostra la distribució dels efectes, els quals se situen entre 0,0055 i 1,88 desviacions típiques, indicant que en tots els casos els resultats del grup d'intervenció són superiors als del grup control.

Figura 5: Distribució d'11 mides de l'efecte



Tanmateix, cal tenir certa precaució a l'hora de llegir aquests resultats, ja que alguns d'aquests estudis tenen limitacions com seria el fet de disposar d'una mostra de participants reduïda o reportar poca informació sobre els continguts de la intervenció avaluada. Al mateix temps, s'identifica que els dos estudis amb mides dels efectes més grans (1,72 i 1,88) fan servir grups de control que en un cas realitza una activitat distractiva enlloc d'una intervenció alternativa i en l'altre rep una intervenció que no persegueix els mateixos objectius d'aprenentatge que després s'avaluen.

No s'ha calculat la mitjana ponderada dels efectes com faria una meta-anàlisi, sinó que s'ha optat per una síntesi narrativa que tracta d'explicar quin tipus de recerca s'està realitzant sobre els efectes dels Tutors Intel·ligents, així com quins factors podrien explicar els efectes descrits en apartats anteriors segons els estudis més recents. A continuació resumim els resultats més destacables.

- **Els efectes d'implementar els Tutors Intel·ligents a gran escala**

La majoria d'estudis primaris identificats són experiments en què s'aleatoritza l'assignació de la intervenció entre un grup reduït d'estudiants. Aquest fet ens pot fer plantejar com de diferents serien els efectes dels Tutors Intel·ligents si el seu ús fos generalitzat, més enllà d'entorns controlats per l'investigador. D'entre els estudis inclosos en la present revisió, el de Pane et al. (2014) de la *RAND Corporation*, destaca per assignar la intervenció a nivell de centres educatius i comptar amb una mostra de 73 centres, corresponent a 18.700 estudiants de *high school*². A més, a diferència de la majoria d'estudis, es tracta d'una intervenció de certa durada, dos anys, i que mesura la fidelitat de la intervenció per assegurar que efectivament s'està implementant tal i com es preveia.

² L'experiment es fa a *middle* i *high school*, però seguint els criteris del protocol de cerca s'ha recollit només la informació de *high schools*.

Per realitzar l'experiment, es van emparellar les escoles participants segons la seva semblança entre elles i es va assignar a cadascuna d'elles la condició d'intervenció o control aleatòriament. La intervenció consisteix en la introducció del *Cognitive Tutor Algebra I* (CTAI) en el currículum del centre. Els resultats de l'estudi indiquen que no hi van haver millores després d'un any de la seva introducció però sí després del segon any. L'efecte és estadísticament significatiu i la seva mida (ES) és de 0,20. Com hem vist en la revisió de les meta-anàlisis existents, es tracta d'un tipus d'estudi que reuneix algunes de les característiques generalment associades a unes mides d'efectes menors: es fa servir un *Cognitive Tutor*, la mostra de participants és gran i l'instrument de mesura dels resultats són tests estandarditzats.

L'estudi és rellevant pel fet de dur-se a terme en un conjunt força variat de centres educatius que no han comptat amb un suport extraordinari per implementar la intervenció. Això demostra que els Tutors Intel·ligents poden ser efectius més enllà d'exercicis puntuals en entorns controlats com serien la majoria dels estudis primaris inclosos en la present revisió.

- **Estudis de VanLehn amb *Dragoon***

En la revisió d'estudis primaris s'han identificat dos articles de VanLehn, l'autor de la primera revisió d'evidència descrita en els apartats anteriors.

En el primer (VanLehn et al., 2016) s'explora si es pot ensenyar com construir models matemàtics en un període de temps curt mitjançant *Dragoon*³, un Tutor Intel·ligent de tipologia *step-based* dissenyat per VanLehn i el seu equip de l'Arizona State University. Per això duen a terme tres experiments que comparen els resultats de classes on s'ha implementat *Dragoon* amb classes que han fet exercicis tradicionals consistents en omplir taules i dibuixar gràfics. En el primer estudi la intervenció es va implementar en tres dies consecutius, en sessions de 50 minuts, i s'observen diferències estadísticament no significatives en els resultats. En el segon experiment la implementació es fa en sis dies, i els resultats mostren que el grup que fa servir *Dragoon* va millorar més que l'altre, amb un efecte (d) de 0,47. Finalment en el tercer experiment dues classes van fer servir *Dragoon* i la tercera classe no, la implementació es fa en tres sessions, i els efectes (d) en la puntuació final són de 0,49 pel que fa a les tasques relacionades a la construcció de mapes conceptuals i de 1 per la resta de tasques.

El segon article de VanLehn és una avaluació dels efectes de *Dragoon* en els aprenentatges sobre la construcció de models de sistemes dinàmics. Es tracta d'una avaluació experimental duta a terme en una classe d'universitat durant tres setmanes, en què 34 estudiants voluntaris van ser repartits en dos grups a partir d'un sistema estratificat. Després de ser numerats segons les seves notes en els primers mòduls del curs, els estudiants amb un nombre senar van ser assignats a una condició i els parells a l'altra (un estudiant va ser assignat erròniament a la condició equivocada, resultant en així en un grup experimental de 16 estudiants i un grup control de 18). Els estudiants dels dos grups ateniaven a classe junts, tenien els mateixos deures i feien

³ <https://dragoon.asu.edu/>

servir *Dragoon*, l'única diferència és que els del grup d'intervenció feien servir la versió de *Dragoon* "entrenada" (*coached*), la qual disposa de *feedback* immediat, mentre que el grup control feia servir *Dragoon* en mode edició, tenint accés a altres materials de suport fora de del tutor (diapositives de *Power Point*). Per comparar els resultats d'ambdós grups es van fer servir les puntuacions en l'examen del mòdul. Els exàmens van ser corregits pel professor de l'assignatura, el qual desconeixia a quin grup estava assignat cada estudiant. La mitjana del grup d'intervenció (41,5) no difereix de forma estadísticament significativa de la mitjana del grup control (38,9).

Aquests quatre estudis contenen algunes de les limitacions comunes en bona part dels estudis primaris inclosos en la present revisió, com la breu durada de la intervenció, la mida reduïda de les mostres de participants, o el fet de que no queda del tot clar el mecanisme d'assignació que es va fer servir, ni si és aleatori o no ho és (en el cas dels tres primers estudis descrits).

- **Integració en entorns d'e-learning**

De tots els estudis inclosos en la revisió, tant sols un ha estat realitzat a l'Estat espanyol, concretament a la Universitat de Valladolid (Verdú et al., 2017) en el marc d'un curs sobre disseny de xarxes. Es tracta de l'avaluació d'INTUITEL⁴, un Tutor Intel·ligent desenvolupat pels autors de l'estudi i pensat per ser implementat en entorns d'e-learning. El sistema proveeix a l'estudiant un guiatge no intrusiu a través de recomanacions, tenint en compte el perfil i comportament de l'estudiant, el context i l'estratègia pedagògica preferida pels professors, els quals poden parametritzar i adaptar el curs a partir de la seva experiència.

Un dels reptes que pot plantejar la implementació d'un Tutor Intel·ligent en un entorn online és la seva integració amb les eines on-line, com podria ser *Moodle* o altres sistemes de gestió de l'ensenyament. INTUITEL ha estat desenvolupat per integrar-se amb aquest tipus d'eines afegint-los un grau de personalització.

El sistema va ser testejat en una sessió per explorar com els estudiants el percebien i hi interaccionaven. Per això es va aleatoritzar l'assignació de la intervenció, creant un grup d'intervenció (10 estudiants), que faria ús d'INTUITEL, i un grup control (nou estudiants), que faria ús de *Moodle* sense INTUITEL. Totes les interaccions dels estudiants amb el sistema van ser enregistrades i després de la sessió es va fer respondre un qüestionari als estudiants sobre el valor percebut d'INTUITEL. Els autors reporten que els integrants del grup d'intervenció van seguir seqüències més similars a allò que havia modelat el professor, tanmateix l'estudi no mesura i compara els aprenentatges d'ambdós grups.

Anàlisi dels estudis que comparen variacions

Els estudis que comparen diferents versions d'un tutor Intel·ligent poden ser molt informatius per entendre com funciona l'aprenentatge amb aquest tipus de sistemes i què fa que puguin ser efectius. Així, els estudis inclosos tracten de respondre

⁴ <http://www.intuitel.eu/>

preguntes sobre com és millor presentar la informació a l'estudiant, quins exemples li poden anar millor per assolir els objectius d'aprenentatge, quines pistes li són més útils per resoldre problemes o quin tipus de feedback el pot motivar a continuar estudiant. A continuació descrivim breument els estudis més destacables.

- **La importància de l'aprenentatge autoregulat**

Un dels conceptes que apareix en freqüència en la literatura sobre Tutors Intel·ligents i altres sistemes adaptatius és el d'auto aprenentatge o aprenentatge autoregulat. Es tracta de l'aprenentatge guiat per la metacognició (el pensament sobre el propi pensament), incloent qüestions com l'avaluació, la planificació i monitoratge del propi progrés.

Barrus (2013), parteix de la hipòtesi de que els estudiants no aconsegueixen els resultats esperats amb els Tutors Intel·ligents degut a les dificultats de transitar d'un model altament estructurat i liderat pel professor, a un entorn altament individualitzat. Per això considera que és necessari un major control sobre el propi aprenentatge. Barrus compara els resultats de diferents versions d'un curs d'àlgebra, en un entorn de *High School*, sobre l'autoregulació, els aprenentatges en matemàtiques i la motivació de l'estudiant. Per fer-ho, compta amb tres condicions: (a) l'ús d'un Tutor Intel·ligent a l'aula, (b) l'ús del Tutor Intel·ligent que incorpora mòduls d'aprenentatge autoregulat, i (c) el grup de control, que no inclou l'ús d'un Tutor Intel·ligent ni continguts relatius a l'aprenentatge autoregulat. Els mòduls específics d'aprenentatge autoregulat tenen objectius com que l'estudiant identifiqui els seus patrons d'aprenentatge, planifiqui la seva pràctica a classe i a casa, es marqui fites o visualitzi els seus objectius. Com a resultat es reporten efectes no significatius. Cal tenir en compte que l'estudi té limitacions importants: La mostra inicial pateix un desgast important, de manera que es perden les observacions d'un bon nombre dels estudiants (gairebé la meitat), i l'assignació de la intervenció no va ser aleatòria, la qual cosa compromet la comparabilitat dels grups.

- **Criteri de mestratge**

Una característica habitual dels Tutors Intel·ligents són les restriccions en la possibilitat d'avançar a una tasca fins que no s'ha assolit el nivell de domini desitjat en la tasca anterior. Es tracta del conegut com a criteri de mestratge (*Mastery criterion*), el mateix que Bloom (1984) i els seus deixebles van testar juntament amb el tutoratge humà individualitzat i als que junts atribuïen una millora de dues desviacions típiques. Paiva, Ferreira i Frade (2017) realitzen un experiment en una assignatura de Màrqueting a l'Escola Superior de Tecnologia i Gestió de Leiria (ESTG), a Portugal, en el qual comparen la implementació d'un Tutor Intel·ligent amb i sense les restriccions del criteri de mestratge. Per això aleatoritzen l'assignació de la intervenció entre 72 estudiants. En el grup d'intervenció l'estudiant només podia avançar si aconseguia una nota igual o superior al 60% en el darrer test. Si això no era així, el sistema suggeria a l'estudiant que realitzés altres tasques o realitzés un altre test similar. Per contra, els estudiants del grup control no tenien cap restricció en l'accés al contingut de l'assignatura però rebien les mateixes recomanacions que la resta. Els resultats de

l'estudi mostren una millora significativa dels resultats gràcies a la introducció del criteri de mestratge, concretament es produeix una millora (d) de 0,87. A la vegada, s'observa que gairebé la meitat dels estudiants del grup control han avançat de nivell sense assolir abans el nivell de domini recomanat, i aquests estudiants han obtingut de mitjana puntuacions inferiors a la resta del grup control.

- **Personalització dels continguts**

La personalització de l'ensenyament és una de les grans promeses dels Tutors Intel·ligents respecte a altres sistemes d'instrucció que fan servir l'ordinador com a suport. A la pràctica són moltes les qüestions que els Tutors Intel·ligents tracten de personalitzar, des del l'assignació de tasques fins a la forma i estil de la comunicació amb l'estudiant. Tanmateix, hi ha qüestions que són menys habituals de personalitzar. És el cas de l'adaptació dels continguts per traslladar-los a un context en el que l'estudiant tingui interès.

Walkington (2013) realitza un experiment amb el popular *Cognitive Tutor Algebra*, en el que la meitat dels estudiants han de resoldre problemes d'àlgebra habituals, mentre que l'altra meitat se li plantegen problemes relacionats amb els seus interessos personals, en àrees com esports, música o cinema. Els resultats mostren un impacte positiu en els aprenentatges, especialment per aquells estudiants amb majors dificultats. Segons l'autor, l'ús dels interessos dels estudiants en les tecnologies educatives adaptatives, com els Tutors Intel·ligents, és una estratègia poderosa per guanyar fluïdesa en sistemes de representació abstractes.

- **Variacions en el feedback que rep l'estudiant**

El *feedback* basat en les experiència i estat de l'estudiant és una característica que diferencia els Tutors Intel·ligents d'altres sistemes menys sofisticats. Es tracta, a la vegada, d'una funció fonamental del tutoratge i com a tal rep importants dosis d'atenció per part de teòrics i desenvolupadors.

Hull i Benedict (2015) analitzen els efectes de proveir feedback motivacional i metacognitiu a través d'un *SQL-Tutor*. Els autors argumenten que la motivació de l'estudiant està estretament lligada amb la seva autoeficàcia, les estratègies d'aprenentatge auto-regulat i els seus aprenentatges. Per això desenvolupen una versió de l'*SQL-Tutor* que incorpora aquest tipus de feedback, a la vegada que demana a l'estudiant que indiqui el seu nivell d'autoeficàcia percebuda a l'inici de cada sessió. La comparació entre els grups d'intervenció i control mostra certes diferències en els aprenentatges a favor del grup d'intervenció, tanmateix són estadísticament no significatives. Tenint en compte la mida reduïda de la mostra (36 estudiants), més recerca podria aportar informació per valorar la necessitat de suport motivacional i metacognitiu.

Per la seva part, Mitrovic, Ohlsson i Barrow (2013) exploren els efectes del feedback positiu en lloc del feedback basat en assenyalar els errors de l'estudiant. Els autors argumenten que el feedback positiu pot reduir la incertesa de l'estudiant sobre les

seves passes temptatives però correctes. Consideren que el feedback positiu ha d'informar sobre tres qüestions:

- Allò que fa que l'acció sigui la correcte
- La descripció de l'acció a un nivell conceptual
- Els aspectes que han canviat en l'estat del problema a través de l'acció

En aquest cas l'experiment també es du a terme amb un *SQL-Tutor* i els resultats tampoc no mostren canvis estadísticament significatius en els aprenentatges, però sí que mostren diferències en el temps que els estudiants necessiten per resoldre els problemes, de manera que el grup d'intervenció resol els problemes gairebé amb la meitat de temps que el grup control. Els resultats de l'estudi suggereixen que els Tutors Intel·ligents poden millorar incorporant feedback positiu que complementi el feedback negatiu.

- **L'ús de diferents tipologies d'exemples**

Booth et al. (2013) testegen si la incorporació d'exemples per il·lustrar tant allò que és correcte com el que és incorrecte pot millorar la comprensió de l'estudiant en el camp de l'àlgebra. Per això duen a terme dos experiments en un entorn de *High School* fent servir un *Cognitive Tutor*. En el primer experiment es comparen quatre intervencions: el Tutor Intel·ligent sense exemples, amb exemples correctes, amb exemples incorrectes, i finalment, amb exemples correctes i incorrectes. Els resultats mostren una millora gràcies a la incorporació d'exemples, però no s'observen diferències segons la tipologia d'exemples. El mateix document conté un segon estudi, però s'ha exclòs degut a que no es du a terme en un curs d'educació secundària post-obligatòria.

Per la seva banda, Reed et al. (2013) es pregunten sobre quina és la millor manera de mostrar els exemples. Per això comparen quatre grups que treballen amb un *Cognitive Tutor* que presenta les següents condicions: exemples en taules estàtiques, exemples en gràfics estàtics, exemples en gràfics interactius i absència d'exemples. No s'observen diferències estadísticament significatives entre els aprenentatges dels grups. Tanmateix, s'observa que els estudiants en el grup control, als qual no s'oferien exemples, demanaven menys pistes i feien menys errors però necessitaven més temps per solucionar els problemes.

- **Diferències en els efectes segons la personalitat de l'estudiant**

Un altre dels reptes i potencialitats dels sistemes adaptatius com els Tutors Intel·ligents és ajustar les seves estratègies a la personalitat de l'estudiant. Al-Dujaily, Kim & Ryu (2013) realitzen dos experiments partint de la hipòtesi de que un sistema particular té efectes diferents sobre l'aprenentatge segons el tipus de personalitat de l'estudiant. En el primer experiment comparen quatre grups formats per la combinació de dues condicions: l'ús d'un Tutor Intel·ligent (Tutor Intel·ligent vs. entorn web no adaptatiu) i la personalitat de l'estudiant (Introvertit vs. Extravertit). L'experiment mostra com en l'entorn web no hi ha diferència entre el resultat segons la personalitat de l'estudiant, però en l'entorn del Tutor Intel·ligent els estudiants extravertits aconseguen més respostes correctes en menys temps. En un segon estudi, en el

mateix document, els autors tracten d'explorar si els resultats del primer experiment podrien ser aplicables a altres tipologies de personalitat. Així, s'agrupen els estudiants en vuit grups segons la seva tipologia de personalitat, però aquest cop tots fan servir un Tutor Intel·ligent i es comparen els seus resultats en coneixement declaratiu i procedimental. Els resultats mostren que:

- Les personalitats extravertides treuen millors resultats en coneixement de tipus declaratiu que les persones introvertides.
- Les personalitats de tipus intuïtiu treuen millors resultats en coneixement procedimental que les personalitats sensorials (*sensing*).
- Les personalitats que tendeixen al pensament (*thinking*) treuen millors resultats procedimentals però pitjors declaratius que les persones que tendeixen a les sensacions (*feeling*).
- No es detecten diferències entre personalitats que tendeixen més a jutjar o a percebre.

Els autors conclouen que la personalitat de l'estudiant influeix en com aquest aprèn en entorns adaptatius i que per tant cal tenir cura de les necessitats de cada tipus de personalitat en el moment de dissenyar sistemes d'aprenentatge.

CONCLUSIONS

La present revisió ens mostra que el volum d'evidència acumulada sobre els efectes dels Tutors Intel·ligents és extens en els àmbits de l'educació secundària post-obligatòria i l'ensenyament superior. Aquesta evidència indica que els Tutors Intel·ligents augmenten els aprenentatges de l'estudiant per sobre del nivell dels mètodes convencionals a excepció del tutoratge humà individualitzat.

Malgrat aquest avantatge respecte altres modes d'instrucció, això no vol dir que seria desitjable que els Tutors Intel·ligents substituïssin tots els mètodes d'instrucció tradicional, ja que l'evidència no assegura que els resultats positius siguin generalitzables a nous contextos i situacions. Tanmateix, sabem del cert que existeixen situacions en tots els nivells educatius i àmbits del coneixement en què els Tutors Intel·ligents poden competir, complementar o substituir de forma exitosa altres modes d'instrucció. És per això que l'ús dels Tutors Intel·ligents hauria de ser una opció a tenir en consideració per part de centres educatius i universitats.

Per altra banda, en el procés de dissenyar i implementar intervencions amb Tutors Intel·ligents hi ha un bon nombre de decisions que cal prendre i per les quals hi ha preguntes que romanen sense una resposta clara. Si bé sembla evident que cal tenir en compte factors cognitius, afectius i metacognitius, es fa necessari un esforç constant i iteratiu per anar ajustant i millorant l'efectivitat de les intervencions, tenint en compte no només els aprenentatges dels estudiants sinó també altres objectius com podrien ser la seva satisfacció i continuïtat en els estudis.

ANNEX 1: NOTES METODOLÒGIQUES

En aquest apartat es presenta la definició d'alguns conceptes clau en la interpretació de la present revisió d'evidència.

Mida de l'efecte

La mida dels efectes (*effect size*) és una manera de quantificar la diferència entre dos grups expressada en desviacions típiques:

$$\text{Mida de l'efecte} = \frac{\text{Mitjana del grup d'intervenció} - \text{Mitjana del grup control}}{\text{Desviació típica}}$$

És especialment útil per calcular l'efectivitat d'una intervenció particular en relació a una alternativa de comparació, estimant no només si la intervenció ha provocat canvis sinó també la magnitud d'aquests. Una mida d'efecte positiva indica que els participants que han fet servir Tutors Intel·ligents han obtingut millors resultats que els que han estat exposats a altres tipus d'instrucció, mentre que una mida d'efecte negativa indica que els participants que han estat exposats a altres formes d'instrucció han obtingut millors resultats que els que han fet servir Tutors Intel·ligents.

Malgrat aquesta simple definició hi ha diferents maneres d'estimar la mida dels efectes segons com calculem la desviació típica que fem servir. Així, l'ES de Glass expressa l'efecte en desviacions típiques del grup control, mentre que la g de Hedge i la d de Cohen ho fan segons la desviació típica combinada d'ambdós grups.

En el cas de la g de Hedge s'afegeix un factor correctiu al càlcul de la desviació típica combinada per evitar el biaix positiu quan la mostra és petita. Quan la mostra és gran la diferència entre les dues fórmules tendeix a desaparèixer.

$$\text{Desviació típica combinada segons la d de Cohen (s)} = \sqrt{\frac{(n_1) * s_1^2 - (n_2) * s_2^2}{n_1 + n_2 - 2}}$$

$$\text{Desviació típica combinada segons la g de Hedge (s*)} = \sqrt{\frac{(n_1 - 1) * s_1^2 - (n_2 - 1) * s_2^2}{n_1 + n_2 - 2}}$$

No tots els estudis primaris reporten les mides dels efectes. En la present revisió, quan no era el cas s'ha calculat la d de Cohen a partir de la informació disponible.

Interpretació de les mides dels efectes

La mida de l'efecte és una mesura estandarditzada expressada en desviacions típiques. Això permet comparar els resultats entre diferents estudis. Tanmateix, el fet de ser una mesura expressada en desviacions típiques pot dificultar la seva interpretació. Per això alguns autors han proposat diferents criteris d'interpretació. En el present estudi agafem com a referència la interpretació de Cohen:

Mida de l'efecte (d)	Interpretació de l'efecte
0,20	Petit
0,50	Mitjà
0,80	Gran

Models per combinar els resultats de diferents estudis

El principal objectiu de les meta-anàlisis és combinar els efectes de diferents estudis per estimar els efectes d'una determinada intervenció. Per aconseguir aquest objectiu, es podria fer la mitjana de les mides dels efectes de tots els estudis inclosos en una revisió, tanmateix no estariem tenint en compte que els estudis tenen diferents graus de precisió. Així, una meta-anàlisi assigna pesos als diferents estudis segons la seva precisió per calcular la seva mitjana ponderada. Per fer això comptem amb dos models:

- **Model d'efectes fixes:** S'assumeix que tots els estudis inclosos són iguals, és a dir, que estudien la mateixa població, la intervenció estudiada no difereix, l'*outcome* d'interès és el mateix, etc. Així, s'assumeix que tots els estudis estan estimant un únic efecte real, i s'atribueix la variació entre estudis a l'error de mostreig aleatori.
- **Model d'efectes aleatoris:** S'assumeix que cada estudi és diferent, i per tant està estimant un efecte específic. Les diferències entre els efectes estimats s'atribueixen a l'heterogeneïtat dels efectes reals de les diferents intervencions avaluades, a més de l'error de mostreig aleatori de cada estudi.

Per fer les ponderacions, el model d'efectes fixes fa servir els errors estàndards de les estimacions dels efectes, la qual cosa és semblant a fer servir la mida de la mostra de cada estudi. En canvi, el model d'efectes aleatoris fa servir un sistema més complex per ponderar els efectes incorporant la variabilitat deguda a diferències entre els estudis.

En el present estudi s'han recollit els resultats de diverses meta-anàlisis. S'ha optat per recollir preferentment els efectes estimats en el model d'efectes aleatoris sempre que aquests estaven disponibles.

BIBLIOGRAFIA

* Al-Dujaily, Kim & Ryu (2013) "Am I Extravert or Introvert? Considering the Personality Effect Toward e-Learning System". *Educational Technology & Society*, Vol. 16 no 3 pp. 14-17

Anania. (1981). "The effects of quality of instruction on the cognitive and affective learning of students".

* Barrus (2013) "Does Self-Regulated Learning-Skills Training Improve High-School Students' Self-Regulation, Math Achievement, and Motivation While Using an Intelligent Tutor?". Arizona State University.

Bloom. (1984). "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring". *Educational Researcher*, Vol. 13, no 6, pp. 4–16

* Booth et al. (2013) "Using Example Problems to Improve Student Learning in Algebra: Differentiating between Correct and Incorrect Examples". *Learning and Instruction*, 25, pp. 24-34.

Burke, A. J. (1983). "Student's potential for learning contrasted under tutorial and group approaches to instruction".

* Butcher, & Alevan (2013) "Using Student Interactions to Foster Rule-Diagram Mapping during Problem Solving in an Intelligent Tutoring System". *Journal of Educational Psychology*, Vol. 105, No. 4, pp. 988–1009

* Chadli, Bendella & Tranvouez (2015) "A Two-Stage Multi-Agent Based Assessment Approach to Enhance Students' Learning Motivation through Negotiated Skills Assessment". *Educational Technology & Society*, Vol. 18 no 2 pp. 140-152.

* Choi (2016) "Efficacy of an ICALL Tutoring System and Process-Oriented Corrective Feedback". *Computer Assisted Language Learning*, Vol. 29 no 2 pp. 334-364.

* Cowan, Choo & Lee (2014). "ICALL for Improving Korean L2 Writers' Ability to Edit Grammatical Errors". *Language Learning & Technology*, Vol. 18, No. 3, pp. 193–207.

* Dzikovska et al. (2014) "BEETLE II: Deep Natural Language Understanding and Automatic Feedback Generation for Intelligent Tutoring in Basic Electricity and Electronics". *Int J Artif Intell Educ* 24, pp.284–332

Graesser, Conley, & Olney, A. M. (2012). "Intelligent tutoring systems". In S. Graham, & K. Harris (Eds.), *APA Educational Psychology Handbook: Vol. 3.* (pp. 451-473).

* Girault & d'Ham (2014). "Scaffolding a Complex Task of Experimental Design in Chemistry with a Computer Environment". *J Sci Educ Technol*, 23, pp. 514–526.

* Grivokostopoulou & Hatzilygeroudis (2013) "Teaching AI Search Algorithms in a Web-Based Educational System". *IADIS International Conference e-Learning 2013.*

- * Hooshyar et al. (2015) "A Flowchart-Based Intelligent Tutoring System for Improving Problem-Solving Skills of Novice Programmers". *Journal of Computer Assisted Learning*, 31, pp. 345–361
 - * Hull & Benedict (2015) "Motivational and metacognitive feedback in SQL-Tutor*". *Computer Science Education*, 25:2, pp. 238-256
 - * Hwang, Chen & Huang (2016) "Development and Effectiveness Analysis of a Personalized Ubiquitous Multi-Device Certification Tutoring System Based on Bloom's Taxonomy of Educational Objectives". *Educational Technology & Society*, 19, 1, pp. 223–236.
 - * Kulik & Fletcher (2016) "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review". *Review of Educational Research*. Vol. 86, No. 1, pp. 42–78.
 - * Lee et al. (2013) "Exploring Optimal Conditions of Instructional Guidance in an Algebra Tutor". SREE Fall 2013 Conference Abstract Template.
 - * Ma et al. (2014) "Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis". *Journal of Educational Psychology*, Vol. 106, No. 4, pp. 901-918
 - * Mitrovic, Ohlsson & Barrow (2013) "The Effect of Positive Feedback in a Constraint-Based Intelligent Tutoring System". *Computers & Education*, 60, pp. 264–272
 - * Mørch et al. (2017) "EssayCritic: Writing to Learn with a Knowledge-Based Design Critiquing System". *Educational Technology & Society*, Vol. 20 no 2 pp. 213-223.
 - * Paiva, Ferreira & Frade (2017) "Intelligent Tutorial System Based on Personalized System of Instruction to Teach or Remind Mathematical Concepts". *Journal of Computer Assisted Learning*, Vol 33 No 4 pp. 370-381.
 - * Pane et al. (2014) "Effectiveness of Cognitive Tutor Algebra I at Scale". *Educational Evaluation and Policy Analysis*. Vol. 36, No. 2, pp. 127–144
 - * Paquette et al. (2015) "Designing a Knowledge Representation Approach for the Generation of Pedagogical Interventions by MTTs". *Int J Artif Intell Educ*, 25, pp. 118–156.
 - * Reed et al. (2013) "Effect of Worked Examples and Cognitive Tutor Training on Constructing Equations". *Instr Sci*, 41, pp. 1–24.
- Steenbergen-Hu & Cooper. (2013). "A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on K-12 Students' Mathematical Learning". *Journal of Educational Psychology*, Vol. 105. No. 4. Pp. 970-987
- * Steenbergen-Hu & Cooper (2014) "A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning". *Journal of Educational Psychology*, Vol. 106, No. 2, pp. 331-347
 - * Stott & Hattingh (2015) "Conceptual Tutoring Software for Promoting Deep Learning: A Case Study". *Educational Technology & Society*, Vol. 18 no 2 pp. 179-194.

VanLehn. (2006) "The behavior of tutoring systems". International Journal of Artificial Intelligence in Education. Vol. 16, 3 pp. 227-265.

* VanLehn et al. (2016) "Learning Science by Constructing Models: Can Dragoon Increase Learning without Increasing the Time Required?" International Journal of Artificial Intelligence in Education, Vol. 26, No. 4, p. 1033-1068.

* VanLehn et al. (2017) "Learning How to Construct Models of Dynamic Systems: An Initial Evaluation of the Dragoon Intelligent Tutoring System". IEEE Transactions on Learning Technologies.

* Verdú et al. (2017) "Integration of an Intelligent Tutoring System in a Course of Computer Network Design". Education Tech Research Dev

* Walkington (2013) "Using Adaptive Learning Technologies to Personalize Instruction to Student Interests: The Impact of Relevant Contexts on Performance and Learning Outcomes". Journal of Educational Psychology. Vol. 105. No. 4, pp. 932-945

* What Works Clearinghouse (2013) "Carnegie Learning Curricula and Cognitive Tutor". WWC Intervention Report

Woolf & Cunningham. (1987). Multiple knowledge sources in intelligent teaching systems. IEEE Expert, Vol. 2, pp. 41-54.