Review Article

Michael Kerres*, Katja Buntins

Recommender in AI-enhanced Learning: An Assessment from the Perspective of Instructional Design

https://doi.org/10.1515/edu-2020-0119
received June 1, 2020; accepted July 1, 2020.

Abstract: As tools for AI-enhanced human learning, recommender systems support learners in finding materials and sequencing learning paths. The paper explores how these recommenders improve the learning experience from a perspective of instructional design. It analyzes mechanisms underlying current recommender systems, and it derives concrete examples of how they operate: Recommenders are either expert-, criteria-, behavior-, or profile-based or rely on social comparisons. To verify this classification of five different mechanisms, we analyze a set of current publications on recommenders and find all the identified mechanisms with profile-based approaches as the most common. Social recommenders, though highly attractive in other sectors, reveal some drawbacks in the context of learning. In comparison, expert-based recommendations are easy to implement and often stand out as simple but effective ways for suggesting learning materials and learning paths to learners. They can be combined with other approaches based on social comparisons and individual profiles. The paper points out challenges in studying recommenders for learning and provides suggestions for future research.

Keywords: recommendation; instruction; design; learning paths; learning resources; artificial intelligence.

1 Introduction

Recommender systems for human learning based on artificial intelligence (AI) are a trending topic in research on *Educational Technology* (Roll & Wylie, 2016). Recommenders are already of great importance in various contexts, and they are the foundation of several online services (Kantor, Ricci, Rokach, & Shapira, 2010). Research on recommender systems *in education* has most often been approached from the view of computer science (cf. Manouselis, Drachsler, Vuorikari, Hummel, & Koper, 2011). Therefore, the paper addresses the question of how these developments can be related to pedagogical threads of discussion and how they conform with evidence from the state of research on instructional design.

In the following section, the discussion about learning paths is first rooted in the pedagogical discourse. In this context, the sequencing of instruction is seen as an essential condition for successful learning and as an important professional routine of trained teachers. The topic has motivated many theoretical concepts, practical models, and empirical research to identify sequences that reliably support learning and to detect moderating variables that enable teachers to adapt their strategies of teaching to, e.g., learning objectives or learner characteristics. Against the backdrop of this discussion, the paper then analyzes the mechanisms of current recommenders for learning and how they try to support learners.

2 Early approaches

Curricula define a progression of contents for instruction. A learning unit describes a sequence of activities accomplished by teachers or learners. Trained teachers know how to sequence these activities to ensure successful learning. Research on instruction as a temporal organization has a long history: one might refer to Herbart's (1776-1841) model of "formal stages" of deepening and reflection, which, however, was already subject to criticism at the time regarding whether such

***Corresponding author: Michael Kerres,** Universität Duisburg-Essen, Essen, Nordrhein-Westfalen, Germany,
E-mail: michael.kerres@uni-due.de
Katja Buntins, Universität Duisburg-Essen, Essen, Nordrhein-Westfalen, Germany,

approaches were helpful or hindered creative learning designs and led to rigid lesson plans (cf. Coriand, 2010).

In the tradition of *direct instruction*, the model "Nine Events of Instruction" by Robert Gagné (1916-2002) defines an ideal sequence as a correspondence of teaching and learning activities (Ertmer, DrisCOll, & Wager, 2014; Reiser, 2001). Other influential works in this field are the CDT by David Merrill (1983) or the 4C/ID model by Jeroen van Merrienboer (1997). *Instructional design* approaches most often rely on instructional sequences with a presentation of content, the assignment of learning activities (especially for practice), and the supervision of the learning progress. The sequence of steps is based on the characteristics of learners or the type of learning subject. The positive effect of these structuring elements on learning (especially for learners with little prior knowledge) has been demonstrated in various studies (Kirschner, Sweller, & Clark, 2006).

The sequencing of learning has also been discussed in the development of computer-based learning software. In contrast to a behaviorist approach, in the 1980s *intelligent tutoring systems* were developed which rely on an ongoing diagnosis of competences. Instead of just tracing *wrong* or *right* answers in multiple-choice tests, the diagnoses are based on observing learners' activities when working on a task. The instruction is then adapted to the current level of diagnosed competence: a system can, e.g., recognize that the concept of 'acceleration' (in physics) has been misunderstood; it can distinguish errors of comprehension from errors in calculation and then adapt the learning path accordingly. To do this, it accesses a database of learning objects and searches for appropriate items that are then presented depending on the level of competence. Despite extensive developments over the last decades (Goodyear, 1991; Mandl & Lesgold, 1988; Sleeman & Brown, 1982), practical implementations of this strand of research have hardly become visible; meta-analyses show rather small effects of these solutions (Kulik & Fletcher, 2016; Ma, Adesope, Nesbit, & Liu, 2014; Steenbergen-Hu & Cooper, 2014).

In the 1990s, attempts were made to *automate instructional design*, aiming at generating learning paths *at runtime*. In contrast to *intelligent tutorial systems* (that focus the ongoing diagnosis of competences), sequencing is based on instructional parameters. In this context, learning objectives are classified, e.g., according to concepts (knowledge) or procedures (skills), and they suggest different sequences of instruction. Here, too, learning objects are stored in a database with descriptors relating to the type of learning objective and its instructional function; for example, a presentation or practice with a certain level of difficulty[1]. The learning path is not hard-wired anymore; an algorithm accesses learning objects that have been classified based on existing categories (for example, this learning element deals with topic X-12 at difficulty level 3; it contains a presentation; the document format is video; it is based on the previous knowledge X-11). Eventually, this ambitious approach failed because neither a validated ontology for the classification of learning content and learning activities nor the knowledge base necessary for defining sequencing existed or could be developed (Spector & Ohrazda, 2004; Tennyson, 1995).

3 The emergence of recommender systems in education

Up to the turn of the millennium, the approaches operated without the internet. The world wide web implemented the idea of *hypertext* to present content in a non-linear fashion (unlike in books or films) (cf. Jonassen, 1986). The implementation of complex learning paths with forced or arbitrary branching became possible, and research examined how students coped with the massive amount of available information (Ellis, Ford, & Wood, 1993; Theng & Thimbleby, 1998; Zumbach & Mohraz, 2008). Early *adaptive hypertext systems* tried to limit the number of links displayed by using filters and presets (cf. Brusilovsky, 1996). The concern of being *lost in hyperspace* mostly proved to be unfounded since learners seem to be able to generate a cognitive map of an application whilst operating with the hypertext universe (Valdez, Chignell, & Glenn, 1988). However, the meta-analysis by Chen & Rada (1996) demonstrates that although non-linear hypertext leads to higher learning outcome and completion rates, linear text proves to be more efficient in many cases: on average, linear text with sequential learning paths contributes to faster accomplishment of learning achievements (see also Unz & Hesse, 1999).

We experience a massive increase in available resources on the internet, and the question arises how computer-generated *recommendations* can help students choose learning resources and how learning experiences can be optimized by sequencing learning paths. This

1 Far less demanding, on the other hand, are support systems intended to support the development of learning software, for example, how to apply the "Nine Events of Instruction" as a learning design (cf. Goodyear, 1994; cf. J. M. Spector & Song, 1995; Tennyson, Barron, & Barron, 1995).

research is part of the broader discussion about artificial intelligence in education using data and modelling to analyze and improve learning. Research and Development in this field has been divided into three main topics with differing aims (cf. Lee, Cheung, & . Kwok, 2020; Wong & Li, 2020):

– AI in educational research: explain and predict learning
– Educational Data Mining: optimize educational programs
– Learning Analytics: improve learning processes

Recommender systems are related to all three strands of research on AI in education. Some systems relate to the institutional level, some systems address the individual learners, some systems are stand-alone solutions, and others are embedded in the learning process. Leitner et al. (2017) describe the range of solutions: they document learners' performances, give feedback on learning progress, make prognoses about learning results (drop-out prevention), and provide counseling and adapt learning paths (Baker & Inventado, 2014; Papamitsiou & Economides, 2016; Zawacki-Richter, Marín, Bond, & Gouverneur, 2019).

Drachsler et al. (2015) have analyzed the range of research on recommenders in education. Based on a systematic literature review, they identified 82 empirically tested systems. They relate to:

– suggesting learning content (n=61),
– suggesting a learning activity (n=4),
– suggesting a learning sequence (n=13),
– suggesting learning partners (n=9) and
– predicting learning outcome (n=1).

These data demonstrate to what extent current research on recommenders covers the different fields of research on AI in education.

4 Five mechanisms of recommender systems

In linearly structured learning programs, sequential learning paths are predefined. Thus, the problem of deciding what to do next does not exist for a user. With the abundance of available materials in hypertext environments, the question arises how learning sequences should be implemented. A single best path could be identified, while other links and materials could be masked or hidden. Yet, a different approach has emerged and has become widely followed: *Recommenders* do not aim at *regulating* the learning process; they *suggest* learning paths and leave the decision to the learner: They offer recommendations but they do not enforce a predefined learning path.

Based on the identified solutions of Drachsler et al. (2015), we examined the applied mechanisms from a perspective of *instructional design*. Interestingly, although most of the papers extensively report on the mathematical principles of their recommenders, they provide only meager information on the instructional design and implementation of their solutions: what kind of tasks the learners were engaged in, what the recommendations looked like and how they were embedded in the learning environment. For a more focused discussion about the potential of recommenders for educational solutions, it seems important to re-analyze these systems from the perspective of instructional design. Based on Drachsler et al. (2015), we therefore derived concrete examples to demonstrate how these recommendations (can) look in learning environments. The inspection of these applications revealed *five* distinct mechanisms for recommenders that are explained in the following sections.

4.1 Expert-based recommendation

Experts provide suggestions for further readings, tests, and other materials. Recommendations are static because they do not depend on other variables. However, personalization can be implemented, for example, by allowing a user to define whether recommendations should be displayed at all or whether only certain recommendations should be displayed. In these cases, the quality of the recommendation depends solely on the expertise of the authors. Probably, they will rely on established, frequently used materials that have proven to be *technically correct*.

Yet, these selected references do not consider the possibly large number of alternatives. They remain static and are not be able to take new resources that have been added later into account. We should bear in mind that the decision-making of learners in a learning process is a qualitatively different task than searching for a movie or buying a car: Students do not want to be *surprised*, they are *not* looking for the hidden bargain, and they do not search for the *latest* releases (of, e.g., movies or audio recordings. Learners are seeking well-established, proven, and correct materials. In contrast to experts, they lack the experience to easily evaluate the quality of learning resources.

Most of the following, far more complex approaches might appear similar on the surface. Therefore, regarding the instructional design, curated recommendations authored by experts should be considered as the *baseline* against which the more elaborate mechanisms must assert themselves.

4.2 Behavior-based recommendation

The recommendation is based on the user's *history*, the sequence of the last visited items, which can be compared with ideal learning paths. In this case, ideal sequences have to be identified, and all items have to be described systematically on the basis of descriptors. The system then checks the last visited pages and compares them with an ideal sequence using the descriptors assigned to these pages. In the following example, it might be suggested that the user should first read an in-depth study text and should then continue to work on an exercise.

Course unit 1
X Introduction - Text
- Interview - Video
X Study text, foundation
- Study text, specialization
- Exercise
- Summary

X = already visited

Figure 1: Recommendations based on previous behavior

We should bear in mind that such sequences are often based on plausibility considerations or the experience of teachers but are rarely empirically validated.

4.3 Profile-based recommendation

The recommendation is based on a profile of the user, which may refer to socio-demographic data, knowledge tests, or scales based on questionnaires. Data can also be extracted from behavioral traces, e.g., an "activity level" is calculated from the frequency and duration of a person's visits, a "communication level" is derived from the number of postings and comments. In figure 2, *heuristics* are listed that can be applied in profile-based recommenders.

For a valid recommendation system, these heuristics would have to rely on empirical validation. In the case of

User with low prior knowledge
Successful users always complete an exercise.
Successful users study text first and then the video before the exercise.
Successful users do not necessarily read the summary.

Users with high prior knowledge
Successful users complete the exercise.

Users with slow operating speed
Successful users first study text introduction, then do the exercise.

Users with fast operating speed
Successful users immediately proceed to the exercise.

Figure 2: Examples of profile-based recommendations.

prior knowledge, for example, the amount of validated research can be rated as high (Kalyuga, 2005), whereas, for learning style, the amount of evidence can be questioned (Riener & Willingham, 2010).

4.4 Criteria-based recommendation

The recommendation is based on defined criteria [i,j], which are used to select those items [A, B, C] with the highest match. Learners make a selection based on these criteria and choose the one they find most interesting from a list of suggested items. On the internet, many applications use this mechanism, e.g., when buying a car, searching for travel destinations, or finding a house on a real estate portal. The mechanism works well when queries rely on criteria that can be specified rather precisely, e.g., a price range, a geographical region, or age. It works less well for books, music, or movies which, for example, can be described by genre or year of publication. Yet, these criteria seem weak and generate relatively many *false alarms*.

In the context of learning, we find the following examples for this approach:

– Search for "videos" that explain the function of an "electric motor". Features might be described as: "Subject area: physics", "Subject: electric motor", "Document type: video", "Display type: presentation, abstract".
– Search for "examples" illustrating the "Pythagorean theorem"; features are: "Subject: mathematics", "Subject: Pythagorean theorem", "Type of presentation: presentation, concrete".

On a webpage, this logic might look like this:

What would you like to do next?
• watch a video on the topic
• work on an exercise
• read a summary
• test my knowledge

Figure 3: Criteria-based recommendations.

4.5 Social recommender

Recommendations can be derived from the behavior of other students and a similarity matrix of their choices and ratings. The following variants are typical for social recommenders and can be implemented in a recommender for learning:

- Users who have worked on this item have also visited that item.
- Users who have visited this item have often visited that item afterwards.
- Users who liked this item also liked that item.

The implementation of this mechanism requires large datasets to recognize underlying structures and to calculate functional models. Batmaz et al. (2019) present typical data sets with, e.g., 1 to 100 million data points. In the context of Massive Open Online Courses (MOOCs), larger data sets are available, but these are not comparable and lack various criteria for applying a *machine / deep learning approach*. For example, the data from different courses can rarely be aggregated across courses, since the composition of the courses often are quite different, and their structures can hardly be compared.

Such *collaborative filtering* is interesting when criteria-based filters lead to unsatisfactory results. This is the case for searches where essential features of items are difficult to describe and lack a shared, intersubjective meaning, as in the case of books, movies, or music: criteria such as 'entertaining', 'quiet', 'positive', 'warm', or 'modern' might be important aspects for a listener but are very subjective. This is precisely where social filters based on similarities help without having to rely on descriptors of items. The underlying problem is not to find the *one* best object (like buying a car or real estate) like a needle in a haystack but rather to discover something new ('serendipity'). Here, the *knowledge of the masses* can be helpful.

5 Application of the classification scheme

Our re-analysis of the solutions compiled by Drachsler et al. (2015) has revealed five distinct mechanisms for recommenders. In the next step, we want to investigate how this classification can be applied to a more recent set of research publications.

Method. To estimate which mechanisms are represented in more recent publications and to what extent, we choose an explicit, transparent, and replicable search und synthesis strategy based on a narrative literature review (Gough, Oliver, & Thomas, 2017). This analytical step aims to check if our classification can be successfully applied to a different literature corpus and deliver enough scholarly articles.

The search was done in the *Web of Science* (Core) using the search words 'Recommender' AND 'Education'. The search string 'learning' was explicitly not chosen, as bibliometric analyses showed that it was related to 'machine learning' and 'deep learning', which are beyond the scope of our research question. The mechanisms of the recommenders were then coded according to our classification scheme.

Only articles from the years 2016- 2019 and only studies dealing with feedback mechanisms within the learning process were selected. Papers dealing with study or course selection and with support tools for teachers were also excluded; thus, 76 articles remained.

Sixteen papers were published in journals relating to education or educational technology, while most of the other papers are situated in the context of computer or information sciences. A closer inspection of the 76 papers led to the exclusion of a further 41 articles because they were presenting a review, covered technology alone, or were not related to our analyses. Of the remaining 35 papers, 16 studies were conducted in the field of education (45.7%), while most of the other studies were related to teaching computer or information sciences (n=19, 54.3%).

Results. Table 1 shows that most solutions rely on a profile-based recommender (n=16), followed by behavioral (n=6), and criteria-based (n=5) recommenders. Social (n=3) and expert-based recommenders (n=1) were only encountered in a few cases. Hybrid formats, integrating several approaches, were found in five studies.

As mentioned, several articles in our sample relate to educational decision problems, such as finding a study program or course, allocating learners to groups, or finding study partners, which were beyond our scope. It seems interesting to note that several of these systems are already implemented as rather routine operations, especially at universities, whereas most of the recommenders focused on in our analysis seem to be earlier prototypes.

6 Discussion

All mechanisms from the classification scheme were found in the corpus from our literature review. There was, however, a clear preference for profile-based recommenders. In the light of these results, we will discuss more closely *if* and *how likely* the different mechanisms, summarized in table 2, can contribute to learning from the perspective of instructional design in the following section.

Table 1: Mapping of articles on 'Recommender' AND 'Education'.

Learning Paths	N
• expert-based	1
• criteria-based	5
• behavior-based	6
• profile-based	15
• social comparison	3
• hybrid	5
Choice of study/courses	8
Tools for teachers	5
Technology	5
Reviews	4
unrelated	19
Total	**76**

Social recommenders can be helpful when it comes to interrelated choices, especially when items are difficult to describe intersubjectively validly, by using a predefined set of criteria or free tags. A recommendation is based on two independent events: Whoever chooses A (or gives a positive rating), also chooses B. The recommendation is not based on conceptual considerations but on estimations based on the relationship between the entities. Such an approach has proven to be an effective solution for various educational problems; for example, in the case of a course recommender such as the one used at Stanford University (Koutrika, Bercovitz, Kaliszan, Liou, & Garcia-Molina, 2009) or the University of Berkeley (Pardos, Fan, & Jiang, 2019), for matching students as partners for cooperative learning (Moubayed, Injadat, Shami, & Lutfiyya, 2020), or for providing suggestions on scholarly research papers for students (Porcel, Ching-López, Lefranc, Loia, & Herrera-Viedma, 2018). Other systems help to find resources for teachers in larger databases (Ruiz-Iniesta, Jimenez-Diaz, & Gomez-. Albarran, 2014). However, when Dwivedi & Roshni (2017) say: "With the advent of web-based e-learning systems, a huge amount of educational data is getting generated", the question remains to what extent this data can be used to generate appropriate recommendations, or whether sufficient knowledge is available on how this data can be processed to present helpful recommendations.

For learning paths *within* courses, however, the amount of available data for building a model is often limited. Data from different platforms or courses often cannot easily be aggregated. Due to structural differences, data from courses like 'Algebra' or 'Clinical Psychology' might be difficult to combine. More importantly, the sequencing of learning resources does not refer to two independent items. In the (often analyzed) MovieLens

Table 2: Recommender mechanisms for learning (summary).

based on …	source	input output	requires	Idioms	common examples
experts	suggestions by an author (static links)	A -> B	Expertise	the single best way	Editor's choice
criteria	criteria the learner has entered	[i,j] -> [$A_{i,j}$, $B_{i,j}$, $C_{i,j}$]	descriptions of (all) items	the needle in the haystack	real estate portal
behavior	current behavior - in relation to ideal sequence	A->B->? [A->B->C]	ideal sequence	following the best path	tax form
profile	characteristics of the learner	? $P_{i,j}$ ->[$A_{i,j}$]	personalized fit	everyone is different	partner search
social comparison	similarities to behavior of other learners	A->? [A->B]'	decisions of other learners	the wisdom of the masses	books, music, movies

or Netflix databases, all items (movies) are categorically equivalent: After watching *one* movie, *every other* movie can possibly be chosen, and the user likes to be surprised (serendipity effect). After having finished one learning activity, on the other hand, *not* all learning objects are a possible “next item”, and the user does not necessarily want to be *surprised*. They want to receive a suggestion for the best next item – on exactly *one* topic, within exactly *this* particular learning sequence. In the sequence of a learning unit, recommendations should keep focus on the current topic and might be attuned to the level of expertise of the learner, etc. A hybrid approach, therefore, might combine social recommendations with previously ascribed descriptors – and could be enhanced with profile-based suggestions (if, for example, difficulty scores exist). Such an approach, however, will most definitely increase the costs and complexity of developing the software substantially.

Additionally, many students might have followed a *wrong* track in the past and visited materials that might be considered questionable or simply wrong by experts. For example, some highly valued training videos on YouTube are criticized because they rely on technically wrong concepts (Orús et al., 2016). DeLong et al. (2006) examine how learners value expert ratings vs. social preferences. Thus, the question remains to what extent the behavior of others can contribute to learning. A solution could be a combination of the two sources: first, recommendations by experts are selected, and then, supplementary social information might be added, such as how often resources have been visited before or how they have been rated by peers. In this scenario, social information is only added to support the individual’s choice.

Criteria-based (or content-based) recommenders seem to be more appropriate than social recommenders in many cases. They depend on a set of criteria that has to be applied to all items and must rely on high interrater reliability. Contrary to Jevisko et al. (2017), we would assume that educational criteria are often fuzzy and lack an intersubjectively or even interculturally shared meaning. Criteria grids for describing a car or real estate, for example, can rely on high interrater reliability because vintage or square feet are variables that rely on intersubjectively valid criteria. For instruction, descriptors would include criteria like: learning objectives, definition of subject areas, expected prior knowledge, intended target group, degree of difficulty, type of instruction, the context in which the topic is dealt with, or type of learning activity required. For these variables, we do not have valid taxonomies with commonly shared definitions. Earlier attempts to define standards for the description of learning objects, such as LOM or SCORM, based on a pedagogically relevant set of criteria (cf. Bakhouyi, Dehbi, Lti, & Hajoui, 2017), reached their limits precisely there: in contrast to criteria like type of document [text, audio, video], pedagogical criteria like level of difficulty do not rely on a shared meaning.

As an alternative, *free tagging* based on keywords that do not follow a prescribed taxonomy has emerged on social platforms. The keywords rely rather on a similarity of terms that can be derived from an underlying semantic network (Golder & Huberman, 2006; Hölterhof & Rehm, 2015; Kerres & Heinen, 2014). The descriptions of items (so far) have to be generated by humans and not – as with social recommenders – by the computer. In contrast to social recommenders, the effort required for the description correlates directly with the number of items to be assigned. When selling a car, this work of categorization is transferred to the individual user, who usually only has to describe a few items and who is highly motivated to categorize the car as precisely as possible. Authors (or editors, e.g., from a publisher), on the other hand, may have to deal with several hundred items regularly, involving a considerable and persistent amount of work.

However, if the criteria rely on a shared understanding by “all” learners and authors and have been assigned to all items, a criteria-based recommender may be helpful. Essentially, it presents a list of options when a user searches for “more” items on a page. In a typical learning environment, the number of available and appropriate items is, unfortunately, often limited. Again, the question remains whether it is not sufficient to simply offer curated items as static links – with the disadvantage that new items will, thus, never be included in such a selection.

Behavior-based recommendations generally appear interesting if “optimal” sequences for different learners are known or seem plausible. The system shows items that have not been visited yet and highlights them as interesting items. Technically, such a recommender could be implemented easily. The system simply needs to remember the visited pages of a course unit. This set can then be compared with an ideal path and the pages not yet visited will be recommended.

For recommendations based on individual **profile characteristics,** these ideal paths must rely on empirical evidence or at least a plausible rationale. Some system designs at hand seem surprisingly *daring*; they do not rely on any empirical evidence: Marchal et al. (2018) attempt to suggest recommendations for learning paths from eye movement patterns. El-Bishouty et al. (2019) use recommendations based on a learning style inventory.

From our set of publications, five papers (mainly from the STEM subjects) use data on *learning styles* (according to Felder-Silverman) for recommendations. However, the construct of “learning style” as a personality trait and its empirical basis is highly debated in the learning sciences (cf. Kirschner, 2017). In further eight papers, prior knowledge is used for the recommendation (mostly based on quizzes). These approaches seem to follow a traditional behavioristic learning design, linking learning paths to results from multiple-choice tests.

Behavioral and profile-based recommendations rely on assumptions about the relationship between (a sequence of) learning activities and learning outcomes, i.e., they assume that certain learning activities contribute to certain learning outcomes. However, our analysis of current recommenders reveals that they only sparsely relate to the state of empirical research even if some knowledge from research exists.

Most of the papers analyzed start with the observation that a vast number of resources exists “on the internet” that might be useful for learning. However, in learning environments, we rarely face the problem of being confronted with thousands of learning resources – yet, this is the (mostly implicitly assumed) key basic requirement of the recommender. Finally, a closer look is necessary if recommender systems address real problems of learners: do learners want to follow the *paths of others*, do they want to find *the needle in the haystack*, receive a *tailor-made offer* or *be shown a gold piece*?

The question arises how the different sources for recommendations are perceived by learners: Martin et al. (2016) found that recommendations provided by *teachers* are by far more valued than those by friends, fellow students, or learning partners. We would assume that learners most often want to entrust themselves to an **expert**; they hope that their learning environment is pre-structured according to valid instructional principles, and, at the same time, that the learning environment offers choices for different learning paths. Expert-based recommendations, far easier to implement than all other variants, are highly valued by learners: “It was also important that the filtered material, which was placed into the ‘Recommended’ content section, was agreed by the teacher so students knew that it was the appropriate/sufficient study material and appreciated this model” (Cerna, 2020, S. 127).

7 Future research

Recommenders in AI-enhanced learning should rely on approved and tested models for learning (either from empirical studies or from machine learning). In some cases, however, it seems that complex formalizations try to overcome the weaknesses of the underlying models. Some recommenders are based on sophisticated formalizations with several elements that would need a complex evaluation scheme: “In the context of a learning style based on an Interpretive Structural Model (ISM), an adaptive learning path recommendation system is proposed comprising: (a) Fuzzy Delphi Method, (b) Fuzzy ISM and (c) Kelly Repertory Grid Technology” (Su, 2017).

The empirical evidence of such models often relies on a study that compares one group of learners receiving instruction with recommendations and one control group without recommendations. If users deem the treatment favorable or if the treatment achieves higher learning outcomes, the quality of the recommender seems to be verified. However, with such a study, it is not possible to assess if and which of the recommendations have been meaningful and are superior to randomly selected suggestions: Instead of providing a control group ‘without recommendation’, it would be more appropriate to compare a control group with *randomly assigned* recommendations.

In double-blind clinical trials, control groups are given a *placebo* and compared against the application of a drug. For a personal learning environment, Chatti et al. (2013) compared 16 different algorithms: They found that “the quality of user experience does not correlate with high-recommendation accuracy measured by statistical methods”, thus, demonstrating the difference between user experience and algorithmic logic and proving the importance of comparing different recommendations against randomly assigned recommendations.

From a learning science perspective, most of the recommenders remain a *black box*: The instructional logic remains hidden for learners and teachers alike. They are not able to comprehend how recommendations come about. With a search engine like Google, the underlying algorithms remain a trade secret. For an educational context, however, the algorithms applied should be transparent and modifiable by learners. Particularly complex mechanisms, on the other hand, delimit reflections and adjustments by users. Still more fundamentally, it might even be questioned if recommenders that are not based on sound evidence can be justified ethically.

In a learning context, recommenders follow – at least implicitly – instructional approaches of *guided discovery learning*, which provide structuring elements in an open learning environment. This approach is backed strongly by empirical evidence (cf. Tobias & Duffy, 2009); it ensures that learning is oriented towards achieving a learning objective without falling back to narrow sequencing strategies from behavioristic approaches which might impede the learning experience. Nevertheless, such approaches have to deal with the repeatedly proven finding that learners tend to ignore or even reject help systems, recommendations, or other advice whilst learning (e.g., Clarebout & Elen, 2009). Therefore, ensuring acceptance of such guidance and support systems is of topmost importance when designing (and studying) a recommendation system.

Before designing a recommendation system, it is also necessary to consider what kind of recommendation learners expect, how learners search for learning resources, and how they construct their learning journey in a given context. Technically complex solutions are not necessarily the first choice. A thorough analysis of the target group and further instructional parameters need to be specified to identify which recommendation mechanism to choose.

Recommenders for movies or real estate are based on preference structures. Eventually, the aim is a purchasing decision and here, similarities prove to be good predictors. There are some differences between the single act of purchasing and the ongoing process of actively engaging in learning and educating. Therefore, the question remains how to conceptualize *guidance* in learning: Should recommendations provide proven routes that have been undertaken by others and that have been successful in the past? Or do we understand education as an opportunity for opening *new* horizons beyond established routes of thinking and for providing new experiences that might *irritate* us (Forneck & Springer, 2005; cf. Stojanov, 2012)?

Education cannot only be limited to the training of common knowledge and skills, continuing to follow paths of learning from the past. Such a view on education would be cautious to present *familiar* resources but would choose to deliver surprising paths that confront learners with unfamiliar concepts and views.

A final note: Recommender in AI-enhanced learning promise choice. By offering guidance and supporting the learner these systems try to improve the individual's learning process. But at the same time, recommenders might contribute to eventually weakening the individual's autonomy and self-regulation because of a dependency on external regulation. When designing recommenders, we should therefore consider deliberately how external guidance can be provided while still keeping the learners' independence and self-regulation as the major learning objective – even in AI-enhanced learning.

References

Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson, & B. White (Eds.), *Learning Analytics* (pp. 61–75). https://doi.org/10.1007/978-1-4614-3305-7_4

Bakhouyi, A., Dehbi, R., Lti, M. T., & Hajoui, O. (2017). Evolution of standardization and interoperability on E-learning systems: An overview. *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–8. https://doi.org/10.1109/ITHET.2017.8067789

Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2019). A review on deep learning for recommender systems: Challenges and remedies. *Artificial Intelligence Review*, *52*(1), 1–37. https://doi.org/10.1007/s10462-018-9654-y

Brusilovsky, P. (1996). Adaptive hypermedia: An attempt to analyze and generalize. In P. Brusilovsky, P. Kommers, & N. Streitz (Eds.), *Multimedia, Hypermedia, and Virtual Reality Models, Systems, and Applications* (pp. 288–304). https://doi.org/10.1007/3-540-61282-3_29

Cerna, M. (2020). Modified recommender system model for the utilized eLearning platform. *Journal of Computers in Education*, *7*(1), 105–129. https://doi.org/10.1007/s40692-019-00133-9

Chatti, M. A., Dakova, S., Thus, H., & Schroeder, U. (2013). Tag-based collaborative filtering recommendation in personal learning environments. *IEEE Transactions on Learning Technologies*, *6*(4), 337–349. https://doi.org/10.1109/tlt.2013.23

Chen, C., & Rada, R. (1996). Interacting With Hypertext: A Meta-Analysis of Experimental Studies. *Human–Computer Interaction*, *11*(2), 125–156. https://doi.org/10.1207/s15327051hci1102_2

Clarebout, G., & Elen, J. (2009). Benefits of inserting support devices in electronic learning environments. *Computers in Human Behavior*, *25*(4), 804–810. https://doi.org/10.1016/j.chb.2008.07.006

Coriand, R. (2010). Johann Friedrich Herbart. In *Zeitgemäße Klassiker der Pädagogik* (pp. 115 126). Verlag Ferdinand Schöningh.

DeLong, C., Desikan, P., & Srivastava, J. (2006). USER: User-sensitive expert recommendations for knowledge-dense environments (O. Nasraoui, O. Zaiane, M. Spiliopoulou, B. Mobasher, B. Masand, & P. S. Yu, Eds.). Berlin: Springer.

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of Recommender Systems to Support Learning. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 421–451). https://doi.org/10.1007/978-1-4899-7637-6_12

Dwivedi, S., & Roshni, V. S. K. (2017). Recommender system for big data in education. *2017 5th National Conference on E-Learning E-Learning Technologies (ELELTECH)*, 1–4. https://doi.org/10.1109/ELELTECH.2017.8074993

El-Bishouty, M. M., Aldraiweesh, A., Alturki, U., Tortorella, R., Yang, J., Chang, T.-W., … Kinshuk. (2019). Use of Felder and Silverman learning style model for online course design. *Educational Technology Research and Development*, *67*(1), 161–177. https://doi.org/10.1007/s11423-018-9634-6

Ellis, D., Ford, N., & Wood, F. (1993). Hypertext and learning styles. *The Electronic Library*.

Ertmer, P. A., DrisCOll, M. P., & Wager, W. W. (2014). The legacy of Robert Mills Gagné. *Educational Psychology: A Century of Contributions: A Project of Division 15 (educational Psychology) of the American Psychological Society*, 303.

Forneck, H. J., & Springer, A. (2005). Gestaltet ist nicht geleitet—Lernentwicklungen in professionell strukturierten Lernarchitekturen. In P. Faulstich, H. Forneck, & J. Knoll (Eds.), *Lernwiderstand—Lernumgebung—Lernberatung. Empirische Fundierungen zum selbstgesteuerten Lernen*. Abgerufen von https://www.die-bonn.de/id/8487/about/html/

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, *32*(2), 198–208.

Goodyear, P. (1991). Research on teaching and the design of intelligent tutoring systems. In P. Goodyear (Ed.), *Teaching knowledge and intelligent tutoring*. Norwood, NJ: Ablex.

Goodyear, P. (1994). Foundations for courseware engineering. In R. D. Tennyson (Ed.), *Automating Instructional Design, Development, and Delivery. (NATO ASI Serie F. 119)*. Berlin: Springer.

Gough, D., Oliver, S., & Thomas, J. (2017). *An Introduction to Systematic Reviews* (2nd edition). Los Angeles: Sage Publications Ltd.

Hölterhof, T., & Rehm, M. (2015). *Sharing Resources to Improve (Teaching) Practice–Do Collaborative Tagging System contribute a Network of Practice among Teachers*. Gehalten auf der Gehalten auf der XXXV Sunbelt Conference of the International Network for Social Network Analysis (INSNA), Brighton.

Jevsikova, T., Berniukevicius, A., & Kurilovas, E. (2017). Application of Resource Description Framework to Personalise Learning: Systematic Review and Methodology. *Informatics in Education*, *16*(1), 61–82. https://doi.org/10.15388/infedu.2017.04

Jonassen, D. H. (1986). Hypertext principles for text and courseware design. *Educational psychologist*, *21 (4)CBT9*, 269–292.

Kalyuga, S. (2005). Prior knowledge principle in multimedia learning. *The Cambridge handbook of multimedia learning*, 325–337.

Kantor, P. B., Ricci, F., Rokach, L., & Shapira, B. (2010). *Recommender Systems Handbook* (1., st Edition.). Springer, Berlin.

Kerres, M., & Heinen, R. (2014). Open Educational Resources and Informational Ecosystems:«Edutags» as a connector for open learning. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, *24*, 154–173.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, *106*, 166–171.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry based teaching. *Educational Psychologist*, *41*(2), 75–86.

Koutrika, G., Bercovitz, B., Kaliszan, F., Liou, H., & Garcia-Molina, H. (2009). *Courserank: A closed-community social system through the magnifying glass*. Gehalten auf der Third International AAAI Conference on Weblogs and Social Media.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lee, L.-K., Cheung, S. K. S., & Kwok, L.-F. (2020). Learning analytics: Current trends and innovative practices. *Journal of Computers in Education*, *7*(1), 1–6. https://doi.org/10.1007/s40692-020-00155-8

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning Analytics in Higher Education—A Literature Review. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning* (pp. 1–23). https://doi.org/10.1007/978-3-319-52977-6_1

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. https://doi.org/10.1037/a0037123

Mandl, H., & Lesgold, A. (1988). *Learning issues for intelligent tutoring systems*. Berlin: Springer.

Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender Systems in Technology Enhanced Learning. In Francesco Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 387–415). https://doi.org/10.1007/978-0-387-85820-3_12

Marchal, F., Castagnos, S., & Boyer, A. (2018). First Attempt to Predict User Memory from Gaze Data. *International Journal of Artifical Intelligence Tools*, *27*(6). https://doi.org/10.1142/S021821301850029X

Martin, E., Hernan-Losada, I., & Haya, P. A. (2016). Comparing social factors affecting recommender decisions in online and educational social network. *New Review of Hypermedia and Multimedia*, *22*(1–2), 6–26. https://doi.org/10.1080/13614568.2015.1027966

Merriënboer, J. J. G. van. (1997). *Training complex cognitive skills: A Four-Component Instructional Design model for technical training*. Educational Technology.

Merrill, D. (1983). Applying component display theory to the design of courseware. In C. M. Reigeluth (Ed.), *Instructional Design Theories* (pp. 61–95). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moubayed, A., Injadat, M., Shami, A., & Lutfiyya, H. (2020). Student Engagement Level in e-Learning Environment: Clustering Using K-means. *American Journal of Distance Education*, *34*(2), 137–156. https://doi.org/10.1080/08923647.2020.1696140

Orús, C., Barlés, M. J., Belanche, D., Casaló, L., Fraj, E., & Gurrea, R. (2016). The effects of learner-generated videos for YouTube on learning outcomes and satisfaction. *Computers & Education*, *95*, 254–269. https://doi.org/10.1016/j.compedu.2016.01.007

Papamitsiou, Z., & Economides, A. A. (2016). Learning Analytics for Smart Learning Environments: A Meta-Analysis of Empirical Research Results from 2009 to 2015. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 1–23). https://doi.org/10.1007/978-3-319-17727-4_15-1

Pardos, Z. A., Fan, Z., & Jiang, W. (2019). Connectionist recommendation in the wild: On the utility and scrutability of neural networks for personalized course guidance. *User modeling and user-adapted interaction*, *29*(2, SI), 487–525. https://doi.org/10.1007/s11257-019-09218-7

Porcel, C., Ching-López, A., Lefranc, G., Loia, V., & Herrera-Viedma, E. (2018). Sharing notes: An academic social network based on a personalized fuzzy linguistic recommender system. *Engineering Applications of Artificial Intelligence*, *75*, 1–10. https://doi.org/10.1016/j.engappai.2018.07.007

Reiser, R. A. (2001). A history of instructional design and technology: Part II: A history of instructional design. *Educational Technology Research and Development*, *49*(2), 57–67. https://doi.org/10.1007/BF02504928

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: The magazine of higher learning*, *42*(5), 32–35.

Roll, I., & Wylie, R. (2016). Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Ruiz-Iniesta, A., Jimenez-Diaz, G., & Gomez-Albarran, M. (2014). A Semantically Enriched Context-Aware OER Recommendation Strategy and Its Application to a Computer Science OER Repository. *IEEE Transactions on Education*, *57*(4), 255–260. https://doi.org/10.1109/TE.2014.2309554

Sleeman, D., & Brown, J. S. (1982). *Intelligent tutoring systems*. London: Academic Press.

Spector, J. M., & Song, D. (1995). Automating instructional design advising. In R. D. Tennyson & A. E. Barron (Eds.), *Automating instructional design: Computer-based development and delivery tools (NATO ASI Serie F. 140)* (pp. 377–402). New York: Springer.

Spector, M., & Ohrazda, C. (2004). Automating instructional design: Approaches and limitations. In David H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd edition, pp. 685–699). Mahwah, NJ: Lawrence Erlbaum Associates. (http://www.aect.org/edtech/26.pdf).

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*(2), 331–347. https://doi.org/10.1037/a0034752

Stojanov, K. (2012). The Concept of Bildung and Its Moral Implications. In K. Schneider (Ed.), *Becoming oneself: Dimensions of* "*Bildung*" *and the facilitation of personality development* (pp. 75–88). https://doi.org/10.1007/978-3-531-19156-0_5

Su, C.-H. (2017). Designing and Developing a Novel Hybrid Adaptive Learning Path Recommendation System (ALPRS) for Gamification Mathematics Geometry Course. *Eurasia Journal of Mathematics Science and Technology Education*, *13*(6), 2275–2298. https://doi.org/10.12973/eurasia.2017.01225a

Tennyson, R. D. (1995). Instructional System Development: The fourth generation. In R. D. Tennyson & A. E. Barron (Eds.), *Automating instructional design: Computer-based development and delivery tools (NATO ASI Serie F. 140)* (pp. 33–78). New York: Springer.

Tennyson, R. D., Barron, A. E., & Barron, A. E. (1995). *Automating instructional design: Computer-based development and delivery tools (NATO ASI Serie F. 140)*. New York: Springer.

Theng, Y. L., & Thimbleby, H. W. (1998). Addressing Design and Usability Issues in Hypertext and on the World Wide Web by Re-Examining the Lost in "Hyperspace" Problem. *Journal of Universal Computer Science*, *4*(11), 839–855.

Tobias, S., & Duffy, T. M. (Hrsg.). (2009). *Constructivist Instruction: Success or Failure?* Routledge.

Unz, D. C., & Hesse, F. W. (1999). The use of hypertext for learning. *Journal of Educational Computing Research*, *20*(3), 279–295.

Valdez, F., Chignell, M., & Glenn, B. (1988). *Browsing models for hypermedia databases*. *32*(5), 318–322. SAGE Publications Sage CA: Los Angeles, CA.

Wong, B. T., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, *24*(3), 875–887.