Showing Academic Performance Predictions during Term Planning: Effects on Students' Decisions, Behaviors, and Preferences

Gonzalo Gabriel Méndez
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador
gmendez@espol.edu.ec

Luis Galárraga
INRIA
Rennes, France
luis.galarraga@inria.fr

Katherine Chiluiza
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador
kchilui@espol.edu.ec

ABSTRACT

Course selection is a crucial activity for students as it directly impacts their workload and performance. It is also time-consuming, prone to subjectivity, and often carried out based on incomplete information. This task can, nevertheless, be assisted with computational tools, for instance, by predicting performance based on historical data. We investigate the effects of showing grade predictions to students through an interactive visualization tool. A qualitative study suggests that in the presence of predictions, students may focus too much on maximizing their performance, to the detriment of other factors such as the workload. A follow-up quantitative study explored whether these effects are mitigated by changing how predictions are conveyed. Our observations suggest the presence of a framing effect that induces students to put more effort into course selection when faced with more specific predictions. We discuss these and other findings and outline considerations for designing better data-driven course selection tools.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**.

KEYWORDS

Visual learning analytics, academic performance predictions, framing effects, course selection, course recommendation.

1 INTRODUCTION

Academic advising is a crucial aspect in the mission of any Higher Education Institution (HEI). One of its central components is course

recommendation. This is of paramount importance to students as proper course selection has a direct impact on their academic workload and overall performance [5].

Course recommendation is usually performed by a designated advisor who assists students in selecting the most appropriate courses for their upcoming term. This advice is based on the advisor's knowledge of the academic program and its history, as well as her ability to craft personalized recommendations from that information. The latter factor makes academic advising particularly challenging: Since each student's history and profile is unique, advisors are repeatedly challenged with previously unseen scenarios that require a thorough analysis. The challenge becomes tougher as advising must often occur within a short period of time [32], making course recommendations prone to errors and susceptible to subjective views. For example, an individual's learning experience may likely influence her perception of the difficulty of a given course. As students lack a global view of the study program, they also tend to make decisions based on the *vox populi*.

For all the reasons mentioned above, some efforts aim at assisting academic advising with data-based visualization tools (e.g., [18, 33]). The goal of such tools is not to replace the human advisor but to empower both students and advisors with complementary actionable advice based on a *more objective* view of the students' enrollment alternatives. That view can be based on official information about the study program (e.g., the courses' number of credits, their expected workload), the historical difficulty of the courses, and the student's historical performance.

A key aspect when designing such data-based tools is how to characterize the performance of students. This is due to the fact that the chosen metric may steer the students' attention to specific aspects of their professional instruction. The GPA, for example, is often seen as a key factor for success in the labor-market and there is a cultural tendency to frame college students based on it [65]. For this reason, it is not uncommon for students to try to maximize their GPA regardless of their actual development of knowledge, skills, or understanding [26]. Hence, the GPA has limitations in reflecting a student's academic performance. From a pedagogical perspective, performance metrics that seek to assess and develop the "21st century skills" [56] (critical thinking, collaboration, creativity, long-life learning, etc.), are more desirable. However, metrics of this kind are seldom collected by HEIs in a systematic fashion. Ultimately, the use of a specific performance metric in a data-based tool must observe any availability constraint. That is, it is dependent on the metrics readily available at the HEIs.

Regardless of the selected performance metric, visualization tools are promising for student-oriented course recommendation because

such tools can efficiently convey the multiple facets of a study program. In this line of thought, we present the findings of two studies carried out with iCoRA (**i**nteractive **Co**urse Selection and **R**ecommendation **A**ssistant), a tool that supports students in deciding their upcoming term's enrollment, prior to their planning advising meeting. iCoRA is part of an initial effort to improve the course recommendation process at the Escuela Superior Politécnica del Litoral (ESPOL), a Latin-American university. The tool's recommendations are based on course grade predictions, which are computed by integrating the available information at ESPOL, namely the student's grades and data about the courses such as workload, number of credits, pre-requisites, and historical performance. iCoRA also provides explanations for its predictions.

The two studies conducted with iCoRA required students to compose and decide on a set of courses for their upcoming term. We first conducted a qualitative study that investigated the effects of showing performance predictions on the students' decisions. Here, the grades predicted by iCoRA were presented through a range-based visual representation. We found that in the presence of these predictions, students focused mainly on maximizing the predicted grades, paying less attention to other important factors that may play a role in their term outcome (e.g., the workload). This aligns with the results of previous research on the unintended consequences of exposing students to historical performance data based on the GPA (e.g., [2, 5, 19, 61]). We argue that this type of overreliance effect constitutes an important limitation of making GPA-based predictions.

In a follow-up quantitative study, we then investigated whether the effects observed in our qualitative evaluation could be mitigated **through design**, by changing the visual representation of the predictions. To this end, we modified iCoRA to convey its predictions through eight different visual representations that span a *specific* to *vague* spectrum. This study focused on characterizing not only the students' decisions, but also their decision process and preferences. We found that some visual representations had significant effects on the students' chosen workload and the time they interacted with the tool's explanations for the predicted grades.

This paper contributes empirical evidence on the impact that grade- and GPA-based predictions have on the behavior of students, as well as the role played by the visual representations of those predictions. We discuss our findings in the context of iCoRA and ESPOL, not without arguing the context-related limitations of our design choices, and the identified effects of showing grade predictions to students. The paper also contributes a discussion on the potential ethical concerns that may arise from providing students with GPA-based predictive tools to support their enrollment decisions. Based on all of this, we devise several considerations and potential principles for the design of new effective data-driven tools for course selection and recommendation.

2 RELATED WORK

Course selection and academic performance prediction are often discussed within the realm of Learning Analytics (LA) [44]. In this section, we first review existing student-oriented visualization tools in the LA literature. Since iCoRA's recommendations are based on grade predictions, we then survey studies on the effects that

exposing students to GPA and historical performance information has on their enrollment decisions and behavior. We conclude this section with the state of the art in visualization design choices and how these affect viewers' interpretation of visual representations. We build upon knowledge from these areas to inform the design of iCoRA and the studies we present in this paper.

2.1 Visual Learning Analytics and Tools for Academic Advising

Viera et al. [67] use the term Visual Learning Analytics (VLA) to refer to LA and Educational Data Mining (EDM) techniques that are facilitated through interactive visual interfaces. Defined as "*the use of computational tools and methods for understanding educational phenomena through interactive visualization techniques*" [67, p. 120], this research area lies at the intersection of LA, EDM, and Information Visualization (InfoVis).

In the area of academic advising, LISSA [18] and LADA [33] are notable examples of VLA tools. LISSA uses historical data to predict the probability of graduation of students within the career's expected time. This is used by advisors to plan enrollment of first-year students who have previously failed courses. Using clustering techniques, LADA predicts the probability that a student fails a course.

Both LISSA and LADA target teachers and advisors as their final users. Student-oriented advising tools are less common. One relevant example in this category is KMCD [70], a self-advising system that shows courses for enrollment based on a given curriculum design. CARTA [70] is another course planning tool that provides students with course descriptive information, evaluations of instructors, and grade distributions. iCoRA shares with KMCD and CARTA the goal of making information on historical data of courses available to students. However, in line with known guidelines for student-oriented VLA tools (e.g., [3, 10, 11, 14, 57, 70]), iCoRA resorts to visualization techniques to also provide performance predictions through visual representations.

2.2 Exposing Students to Historical Performance Information

Several research efforts have investigated the impact of disclosing information about performance of previous cohorts on students. According to Ognjanovic et al. [53], the knowledge of historical GPAs is a key factor to explain the courses students opt for. It has been found that when students have access to the performance outcomes of previous courses, they tend to choose leniently graded courses [5] or make shortsighted choices regarding their careers [63]. In a more focalized context, Lim et al. [43] found more recently that even Learning Analytics Dashboards (LADs) may have a negative impact on students because of the social anxiety they experience when their peers performance is compared to theirs. These and other unintended consequences [19] often prevent HEIs from making performance and GPA information publicly available.

On the other hand, a parallel line of research found that when students are indirectly exposed to academic performance visualizations through their advisors or counselors during one-to-one meetings, they show—over a relatively short period of time—positive changes in motivation and self-regulated strategies for learning [2].

Along the same lines, Main & Ost [47] identified that there was no evidence of the effect of letter grades on the students' enrollment decisions. They also found a positive effect on the students' efforts within courses.

The body of work referred above suggests that there is still the need to study the impact of LADs and data-based tools that expose students to historical performance information. We take steps in this direction with a special focus on visualization, by also investigating the role that different visual representations play when presenting performance predictions to students.

2.3 Frames and Visual Representations

A framing effect arises when people make different choices based on how a given problem—or set of options—is presented. This type of cognitive bias has been widely studied in opinion formation (e.g., [22, 25, 52]) and decision making processes (e.g., [40, 41, 48]). Framing effects have also been observed when visual representations are used as communicative structures of a message. Cheema et al. [20] found that visual representations for goal progress (e.g., progress bars) enhance motivation as people approach their goal. Low-level visual features such as spacing, position, and order have also been found to impact the responses elicited by survey questions as well as the response process [66]. Baumer et al. explored how framing effects can be mitigated in text visualizations of political issues [7, 8]. Other explorations in the context of human rights narratives have investigated the effects that anthropomorphizing standard charts has on the empathy and prosocial behavior of the viewers [12].

In a broad sense, rhetorical techniques—the choices made at the data, visual representation, annotation, and interactivity levels—steer our thinking of the topics presented by a visualization. In consequence, those techniques affect end-user interpretation [34]. The way different design choices prompt viewers to interpret visualizations from different perspectives has been investigated at different levels: from the effectiveness of low-level visual mappings [23, 24, 69], to the impact of more high-level concepts and visualization elements such as titles [38, 39] and visual embellishments [6]. The impact of the latter group has been studied in different contexts: visualization recognition, recall, comprehension and interpretation, memorability, perception of bias, and change of attitude. All these are important processes that viewers experience when exposed to visual representations of data.

Inspired by the body of knowledge summarized above, we are interested in investigating the effect that different visual representations of academic performance predictions have on the students' decisions when planning their upcoming terms. In this work, we define a continuum of prediction representations—ranging from *specific* to *vague*—and study how the students' decisions, decision processes, and preferences are shaped by these representations.

3 MOTIVATING CONTEXT AND RESEARCH QUESTIONS

ESPOL[1] is an engineering-oriented Ecuadorian university with over 10,000 students and 32 undergraduate programs. The advisors of its academic advising system are lecturers chosen by workload availability who are assigned up to 40 students (25 on average). Advising sessions take place twice every term over a two-week period: right before the term begins (for course selection and recommendation) and after the midterm exams (to monitor the students' performance). Each advising session is supposed to last no longer than 15 minutes.

In-house observations and interviews revealed that it is common for students to arrive unprepared or undecided to their term planning advising appointments. This makes advising sessions longer, which is particularly problematic when the students have other issues that also need to be addressed during the meeting. Besides, this lack of preparation may induce the students to select their courses on the spot, likely on the basis of unofficial, incomplete, and potentially non-accurate information. For instance, in-house inquiries about the activities students perform to decide on their courses, reveal that 83% ask other fellow students not only about the difficulty of the courses, but also about the reputation of lecturers. These inquiries also indicate that students deem their fellow students' advice as important as their advisor's.

At ESPOL, students pass a course with a minimum grade of 6.00 (out of 10) and are ranked in terms of their GPA, which is reflected in their official academic record and transcripts. Although some instructors may conduct class activities using alternative performance metrics (e.g., development of learning outcomes, levels of engagement), ESPOL's current grading policy enforces all evaluations to be captured via the students' grades and, consequently, their GPA. This information is also commonly requested by recruiters of the local market in job applications. For these reasons, the students at ESPOL deem GPA performance highly important, so much so that they carefully consider any potential impact on their GPA when making enrollment decisions.

The aforementioned observations suggest that a tool that supports the data analysis aspect of course selection could help students not only in preparing for their term planning appointments, but also in making more informed decisions. Ultimately, this could alleviate the advisor's workload and provide students with a more objective view of their study program. We highlight that, rather than replacing the advisor, a tool of this type has the potential to make the student-advisor dialogue more effective and efficient. However, before such a tool could be deployed in a real-world setting, we would need to understand:

RQ1: What are the effects of showing performance predictions to students during term planning?
RQ2: How do these effects vary when we change the visual representations used to convey the predictions?

We investigate these questions through the lens of iCoRA [15], an interactive visualization tool that provides students with historical data on their academic program. The studies conducted with iCoRA focus on the Computer Science (CS) program of ESPOL, that is composed of 41 courses (104 credits). 37 of these courses (96 credits) are compulsory while the remaining 4 (8 credits) are elective. This curriculum design makes the enrollment less flexible

---

[1]http://www.espol.edu.ec

than most universities in Europe and North America, where students can often *mix and match* a wider variety of courses based on their interests and tastes.

Given the extensive use of the GPA at ESPOL and its importance for the students' career prospects, iCoRA's current implementation issues course recommendations based on the students' past grades. That said, we do acknowledge the limited capacity of the GPA to fully describe a student's learning, capabilities, and skills. A myriad of factors beyond course grades have shown to influence students' performance (e.g., demographic and socio-economic background [30]; high school history [37]; social ties with classmates [28]; personality and psychological aspects such as self-efficacy [4], motivation [58, 59], and approaches to learning and preferences for teaching/courses [17]). Therefore, this investigation should be regarded as evidence of the effects of exposing students to performance predictions *in general*. Our goal is to provide a reference for the design of tools based on other performance metrics, by showing how certain design choices may shape the students' behavior (see also section 7.4 in the Discussion).
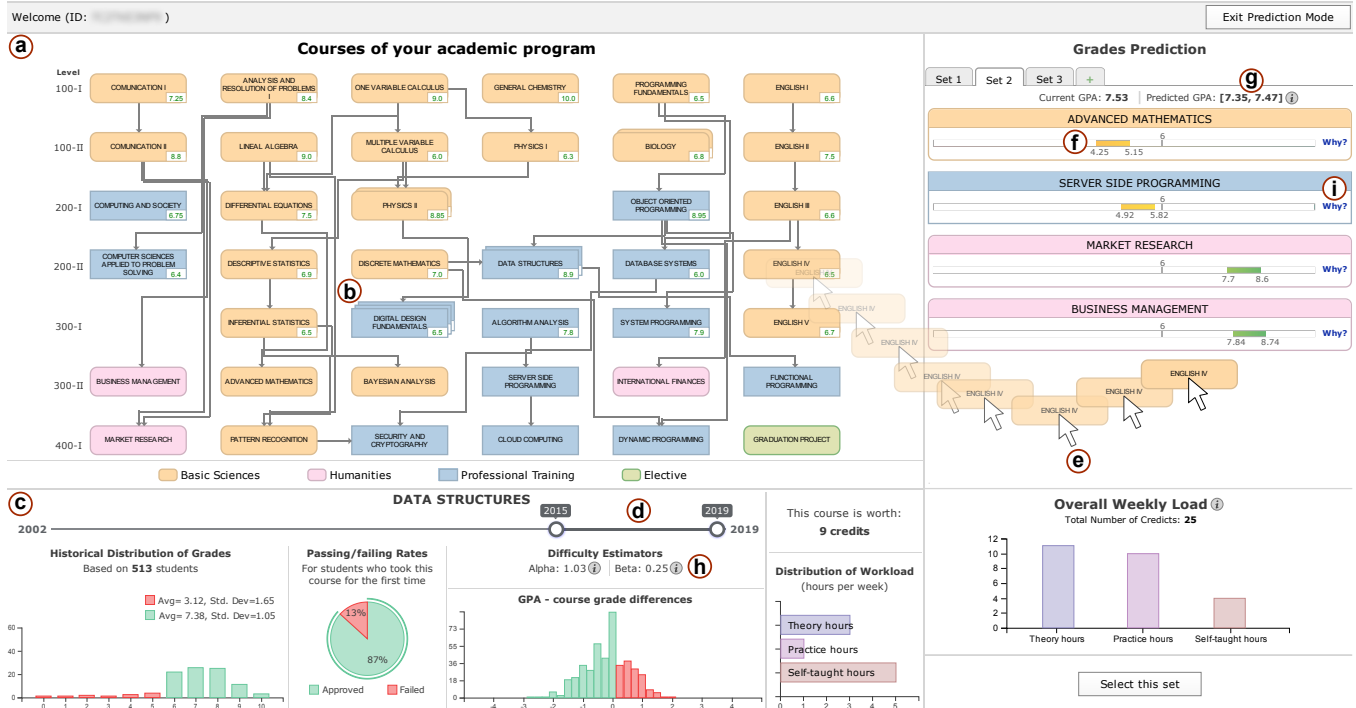
4 ICORA

iCoRA [15] is a tool that assists students in planning their upcoming term in preparation to their advising appointments. It supports the composition of arbitrary sets of courses available for enrollment. Based on past observations, it provides performance predictions and information on the term's resulting workload and difficulty.

Although iCoRA is not the main contribution of this paper, this section describes the tool in detail as its components are relevant for the studies later described.

4.1 Students' Academic Program and History

The *program view* shows the student's academic program as a grid of courses with links indicating pre- and co-requisites (Figure 1a). Courses are organized into four categories (basic science, professional training, humanities, and elective) and are color-coded accordingly. This view shows each course with the grade obtained by the student; the grades are shown in green for passed courses, and in red for failed ones. Courses that have been repeated are depicted as groups of stacked rectangles, each representing an enrollment instance (e.g., Figure 1b).

Clicking on a course of the program view displays the course's general and historical information (Figure 1c): number of credits, weekly workload, difficulty estimators (course grading standard α and grading stringency β—as defined in [16]), distribution of grades, and historical performance. This data can be filtered by time through an interactive range slider (Figure 1d). This supports the exploration of the course's evolution over time and provides insights about the performance of students who have recently enrolled in a given course. This is relevant to support students in making decisions in light of recent data.

Figure 1: iCoRA's student module. The main view shows the courses of the student's academic program. Clicking an element of this view reveals the course' history and general information. Under the prediction mode, courses can be dragged onto the prediction panel. In response to these interactions, iCoRA predicts the student's performance in each of the selected courses.

4.2 Course Sets and Performance Predictions

Under the *prediction mode*, available courses from the program view can be dragged onto the *grades prediction* panel (Figure 1e) to compose one or more sets of courses. These interactions trigger the execution of iCoRA's performance prediction models and update the panel's content.

The prediction models for each subject are based on gradient boosting trees (GBT) trained on historical data that comprise term workload, previous grades, failing history, and aggregated course difficulty. In the version of iCoRA shown in Figure 1, the performance prediction of each course is depicted as a range—computed via quantile regression on GBT—on a horizontal scale between 0 and 10, in compliance with ESPOL's grading system (Figure 1f). The range is shown through a red-yellow-green divergent color scale with a zero value of 6.00—the minimum passing grade.

On adding courses to—and removing them from—the prediction panel, iCoRA estimates the student's GPA that would result if the predictions shown became true. The GPA is estimated by considering the lower and upper bounds of the ranges predicted and is presented on the interface also as a range (Figure 1g).
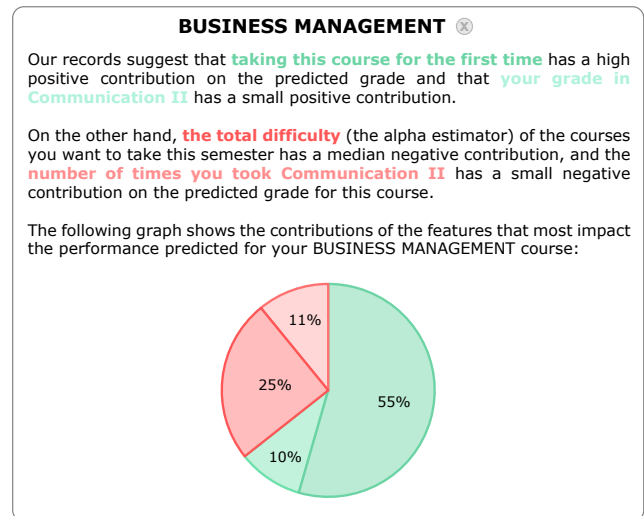
4.3 Explanations

iCoRA provides explanations of some of the features used by its prediction models. These explanations combine text, very simple visualizations, and math formulas. Examples include tooltips describing the difficulty estimators for courses (Figure 1h). The performance predicted for each course is also explained. The *Why?* button to the right of each prediction (Figure 1i) explains the relative contribution of the model's input features to its output (see Figure 2). This contribution is calculated with SHAP [45], an explanation method based on linear feature attribution.

iCoRA offers these explanations so that the students can capitalize on the factors that could positively influence their performance. Perhaps more importantly, these explanations seek to encourage students to mitigate potential negative impact. For example, a way to reduce the risk of getting bad grades could be to decrease the overall grading stringency (total β) of the courses. This could be done by enrolling in fewer courses or by taking easier ones.

Having introduced the functionalities of iCoRA, we are ready to elaborate on the user studies we conducted with the tool to answer our research questions. These studies investigate the impressions of the students with regard to iCoRA's functionalities, and in particular, shed light on the impact of performance predictions on the decisions, behaviors, and preferences of students in the context of course selection.

5 STUDY 1 — SHOWING PERFORMANCE PREDICTIONS

We ran a qualitative study that investigated **the effects that showing performance predictions has on students when they plan their upcoming term (RQ1)**. Our original experimental design was based on a controlled lab study. However, the sanitary crisis around the COVID-19 pandemic forced us to convert our protocol into a remote format. We thus used video conferencing software to test and interview participants remotely.

Figure 2: Explanation of the grade predicted for a *Business Management* course. Besides the pie chart, the version of iCoRA used in our first study included a written summary of the impact of the model's input features.

5.1 Participants and Procedure

We recruited 12 participants from ESPOL' computer science (CS) undergraduate program (4 female; 8 male; 21–30 years old; median age 24). Students were at different stages of their degree: first (n=2), second (n=6), and third (n=4) year. All had attended at least two academic advising meetings.

In each individual study session, participants were asked to put themselves in the place of the fictional student whose academic history and set of available courses were shown in iCoRA. Participants had to select a set of courses for their upcoming semester, and were allowed to use iCoRA to work on this task for as long as they wanted. We did not specify restrictions regarding the number of courses they were allowed to take.

Participants worked with a modified version of ESPOL's CS program, where the last two semesters were replaced with courses from other CS curricula. This was enforced by ethics regulations in order to avoid influencing the students' attitude towards actual courses they had not taken yet. The introduced courses were chosen so that they seemed plausible, that is, they had names that participants could understand and relate to (e.g., *Dynamic Programming*).

Participants had to choose among a set of nine introduced courses distributed across basic sciences, humanities, and professional training. Each category had three courses of low, average, and high difficulty.

For the sake of the study, iCoRA was fed with synthetic data. Grades and aggregated difficulty estimators were randomly drawn from different normal distributions skewed according to the courses' difficulty. A student's failing history (number of times a course was taken) was generated using a power-law distribution. The models that predicted the performance of the surrogate courses consisted of handcrafted linear functions that allowed us to control the contribution of each feature to the predicted performance intervals.

5.2 Data Collection and Analysis

We used online questionnaires to collect participants' consent, demographic information, and data on the strategies they usually follow when choosing their courses. We recorded all the sets of courses composed by the participants. We also captured their interactions with iCoRA through a video conferencing tool. In a post-task questionnaire, participants rated a series of propositions about iCoRA. The interviews were recorded, fully transcribed, and qualitatively coded following a thematic analysis approach [13]. Our initial coding was done by two researchers independently and focused on the students' general perception and rationale. Higher-level topics emerged in subsequent meetings in which the coding scheme was revised iteratively by the two researchers until a unified coding scheme was reached.

5.3 Results

Our analysis of the questionnaires and interviews revealed a general enthusiasm for iCoRA. Participants particularly appreciated the access to their courses' historical information and highlighted the usefulness of this feature to get an overall impression of a course's reputation—instead of having to ask other fellow students about this. The prediction feature was highly appreciated—both at the course and the GPA levels. Figure 3 shows a summary of the participants' ratings of several aspects of iCoRA. These results are displayed in a 7-point Likert scale.

In the subsections that follow, we present the most important findings of this study. The quotes included below have been translated from Spanish.

5.3.1 Students' Decisions. Before the course selection task, we asked participants about the strategies they usually follow to decide on their courses. Their answers included aspects such as aiming at a specific term workload ("*I choose between five and six courses per semester*" [P01]; "*I always choose four courses*" [P11]), balancing the difficulty of their courses ("*I have to choose this course [...] with easier courses*" [P03]), and following the sequence in which courses appear in their academic program ("*I usually don't choose courses from distant levels*" [P01]).

After composing and choosing a course set with iCoRA, during the interviews, we asked students on the rationale behind their decisions. All participants, with no exception, considered the predicted performance of the courses as the most important factor to select their courses: "*I noticed the grades were better in my second set of courses. So, I chose that.*" [P03]; "*[iCoRA] showed me the minimum grade I was going to get and that's important because it affects my GPA for the next semester.*" [P11]. In five occasions, performance was also mentioned in regards to the predicted GPA: "*It showed me how my GPA was going to improve by the end of this semester*" [P07].

These statements suggest that iCoRA's predictions heavily influenced participants' approach to course selection. The tool seemed to have turned the participants' attention to the predicted grades, away from other aspects that students traditionally consider when deciding on their courses. We found that, in the presence of performance predictions, students perceive course selection as a grade maximization problem. The data supports this hypothesis: The set
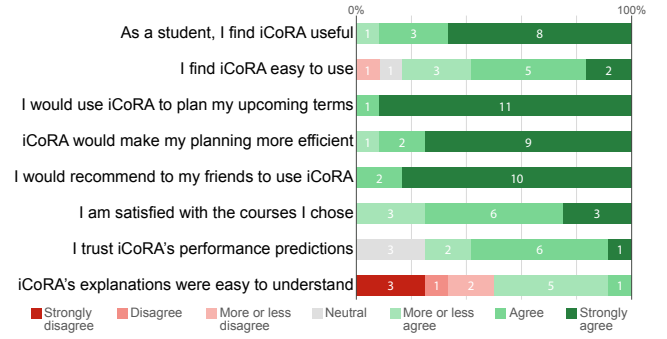
Figure 3: Participants' ratings on several propositions about their experience with iCoRA. Each proposition was rated using a 7-point Likert scale.

of courses selected by the students are, on average, at the 96-th percentile in terms of GPA's predicted upper bound when we consider all the course sets they ever composed. When we look at the GPA's lower bound, and the maximal individual grades, the sets lie at the 77-th and 87-th percentiles respectively.

Our video analysis also indicates that when selecting courses, students often disregarded factors such as the workload they would face or the difficulty of the chosen courses.

5.3.2 Participants' Interest in Explanations. Explainability is crucial to produce predictions that humans can understand and trust. iCoRA takes steps in this direction by providing explanations for the course difficulty indicators α and β [16]. In the same vein, the *Why?* button explains the impact of the model's input features on the prediction outcome. This functionality aims at opening the *black boxes* used by the tool. However, the participants' ratings on the explanations suggest that these might have not been very effective. Some students commented on this explicitly: "*The explanations could be less formal.*" [P05]; "*There are too many words, they could be replaced with icons, or perhaps be more concise.*" [P06]; "*Show them with other words, they were hard to understand.* [P11]."

The comments of above suggest the version of iCoRA used in this study has room for improvement regarding how it explains the different pieces of information—which seemed not obvious to the students. However, our video analysis also revealed that overall participants interacted very little with the explanations. Regarding the difficulty estimators of the courses, only three participants opened the explanation for α (mean time 8 seconds) and just one checked the explanation for β (during 25 seconds). The explanations for the performance predictions provided through the *Why?* button sparked more interest: ten participants opened them at some point, leading to a global average of 1.5 minutes (for the total time). However, the participants' interest in these explanations decreased significantly after their first interaction with them. Only eight participants requested these explanations a second time and the average time they spent on it went from 53 seconds for the first time to just 12 for the second. Further interactions with the *Why?* button were very rare, and always shorter.

6 STUDY 2 — ALTERNATIVE VISUAL REPRESENTATIONS OF PERFORMANCE PREDICTIONS

Motivated by the observations presented above, we designed a second study to better understand whether and, if so, how students are influenced by the way iCoRA's performance predictions are conveyed. The driving research question for this study was **whether different visual representations for performance predictions can affect the decisions and behaviors of the students when they select their courses (RQ2)**. We wanted to see, for example, if text-based performance predictions would make students less eager to maximize their grades. In this study, we investigate these effects not only on the students' final decisions, but also on their decision process and preferences.

We followed a protocol similar to the one of Study 1 but this time, students had to choose courses from several versions of their academic program, each with a different set of available courses. Besides, the performance predictions were displayed using different visual representations. Before describing our experimental protocol in detail, we first explain the alternative visual representations we used to answer RQ2.

6.1 A Spectrum of Performance Prediction Representations

For this study, we designed eight different ways to show performance predictions and integrated them into iCoRA. These representations span along a spectrum from *specific* to *vague* (Figure 4). This spectrum is inspired in work by Walny et al. [68] that describes a continuum of visual representations from countable (numeric) to pictorial (abstract), found by observing how people sketch representations of data. A set of similar representations was found by Méndez et al. [49] after comparing the visualization construction process of iVoLVER [50] and Tableau Desktop. Based on these continua, we consider a performance prediction representation to be more **specific** if it makes the actual grade more directly readable, and more **vague** if it manipulates the grade to represent it graphically, in a more abstract way. We elaborate on the visual representations that compose our spectrum in the following:

- value: Shows the predicted grade with a line mark along a [0.00 − 10.00] horizontal scale. The line mark is colored according to the red-yellow-green scale of the range representation used in Study 1 (Figure 5b).
- range: This is the range representation used in Study 1. It shows the lower and upper bounds of the interval predicted by our models. It uses the same continuous color scale of Study 1 (Figure 5c).

- bars: Fills a portion of a horizontal bar with color indicating the course type (e.g., humanities). The bars of all the selected courses are aligned, which essentially composes a horizontal bar chart with a common left baseline (Figure 5d).
- stars: Represents the predicted course grade by filling a set of five stars—similar to those used in rating systems. This representation could be considered as a discrete version of the bars one. Color is also used here to depict the course type (Figure 5e).
- area: Uses circular marks that scale relative to each other to represent the predicted grades of a set of courses. Color is used here to depict the course type (Figure 5f).
- color: Uses full, single-colored bars to encode the grade of each course. The color comes from the red-yellow-green color scale used by the value and range representations (Figure 5g).
- text: Shows a text qualifying the course's predicted grade. The tone of the message varies from "*It is very likely that you will fail this subject; you will have to prioritize it over your other courses.*" for grades between [0.00 − 2.00] to "*You will do excellent in this subject; your grade may make you look exceptional in relation to other students.*" for grades in the range [9.00 − 10.00] (Figure 5h).
- faces: Shows a colorless emoji-like face made up of two circular eyes and a curved mouth. As the course grade gets closer to 10, the eyes scale up and the curvature of the mouth increases. This is a very minimalistic version of the Chernoff faces [21] (Figure 5i).

These representations aim to cover a wide range of levels of specificity at conveying a predicted grade. The ends of the spectrum represent grades in very different ways: the value representation is very specific, whereas the faces representation requires decoding the grade from an abstract representation. We consider the representations to be split equally between the *specific* and *vague* categories: Four of them (value, range, bars, and stars) are located at the **specific** side of the spectrum while the remaining four (area, color, text, and faces) lie closer to the **vague** end. However, we remark that the exact position of each visual representation along the continuum should not be considered definitive. Especially within each category, some representations have similar effectiveness to encode quantitative values [23, 24, 51]. This is particularly true for the bars and stars representations of the *specific* category and for the area and color representations of the *vague* end.

Figure 5 provides examples of predictions using the representations of our spectrum. The example shows the performance predicted for a set of three courses: *Business Management*, *Advanced Mathematics*, and *Dynamic Programming*. Figure 5a shows how the tool presented this set of selected courses when no prediction was provided. The remaining ones show the *specific* (Figure 5b–e) and *vague* representations (Figure 5f–i).

Figure 4: *Specific* **to** *vague* **spectrum of visual representations for performance predictions. See Section 6.1 for a detailed description of each representation.**
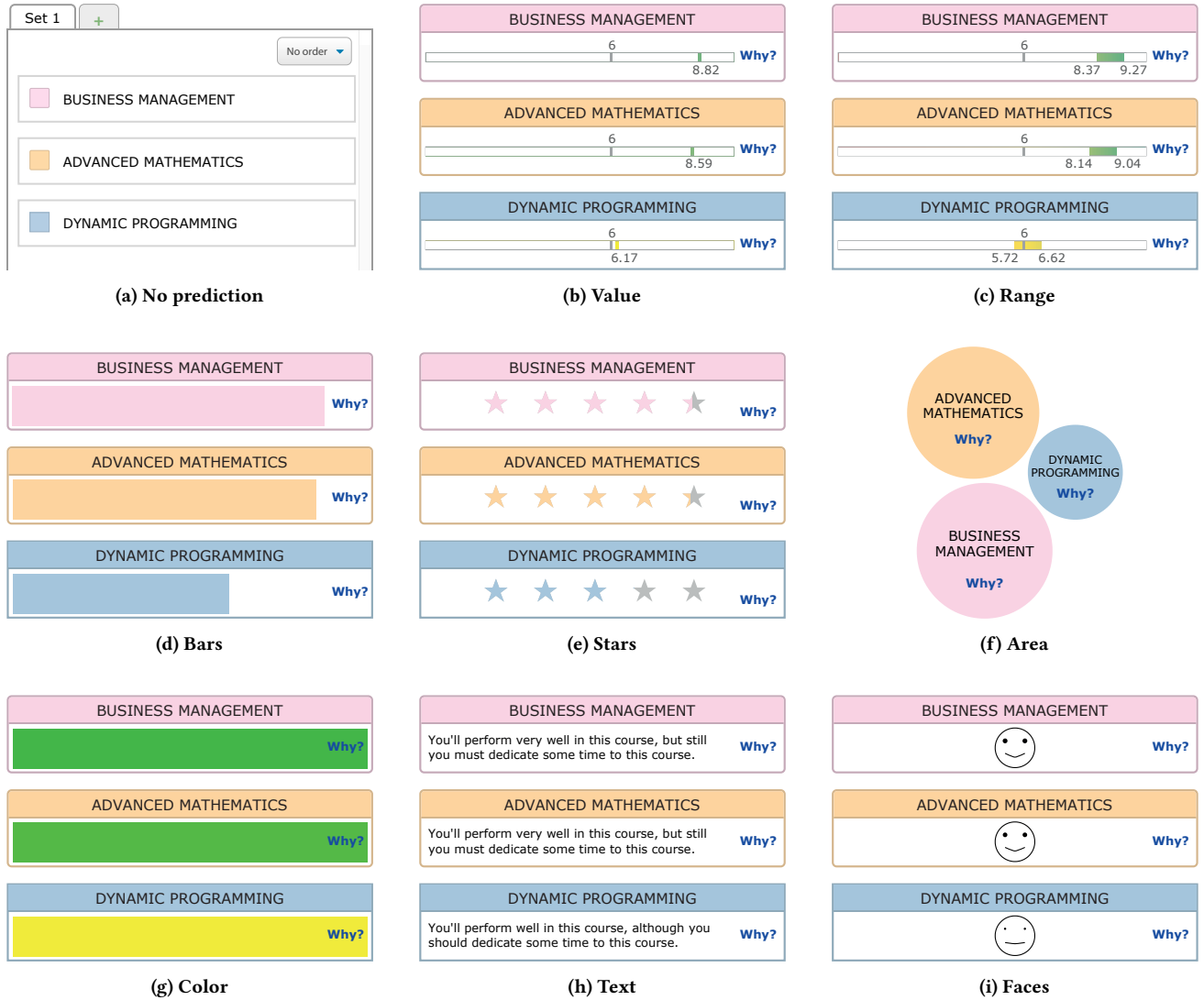
Figure 5: Performance predictions from the spectrum of Figure 4 as show by iCoRA in Study 2. Sub-figure (a) shows how the interface presented a set of selected courses in the control condition (i.e., when no performance prediction was provided).

6.2 Experimental Design

For Study 2, we followed a between-group design with respect to the **prediction representation type** (*specific* and *vague*). Each of these independent variables has four levels: value, range, bars, and stars for the *specific* condition; and area, color, text, and faces for the *vague* one.

For each of these two conditions, we had two dependent variables: student's decisions and behavior. The students' decisions (i.e., the course set they chose) are operationalized via four dimensions: the **number of selected courses**, the **average predicted grade** of those courses, the course set's **average workload** (expressed in hours per week), and its **total workload**. On the other hand, the students' behavior during a course selection task was operationalized through the time they interacted with the explanations for the performance predictions. More specifically, we measured the student's behavior by the **number of times** they invoked the tool's explanations, and the **total time** these explanations remained open.

We then used the levels of the independent variable **prediction representation type** to conduct a within-subject analysis for each experimental condition. These analyses compared the measurements of the dependent variables within each level (*specific* and *vague*).

6.3 Participants

For this study, we invited students of two *Human-Computer Interaction* and one *Data Structures* courses. Out of the 105 students enrolled in these courses, 91 volunteered to participate (74 male, 17 female; 19–32 years old—median 22). All were enrolled in ESPOL's CS undergraduate program and none had participated in Study 1.

All have had prior academic advising and were at different stages of their degree: second or third year (n = 39), fourth (n = 27), and superior years (n = 25).

6.4 Procedure

We modified the version of iCoRA used in our first study to display the sequence of forms, tasks, and questionnaires that participants had to work with. We made this modified version of the tool available online. It had a wizard-like interface design that guided participants through the following sequence of activities:

Introduction to iCoRA. After providing consent and filling out a questionnaire about their demographics and course selection habits, each participant watched a 12-minute video that explained iCoRA's user interface. The video described how to compose sets of courses, and the tool's performance predictions and explanations. It also elaborated on the tasks participants had to complete.

First course selection task: No Prediction mode. In this study participants had to complete five course selection tasks, always starting with a scenario in which iCoRA did not display any performance prediction (as shown in Figure 5a). Similar to the procedure of Study 1, participants were instructed to put themselves in the shoes of the student whose academic history and set of available courses were presented. They had to compose a set of courses to enroll in *their* upcoming term and submit their selection.

We introduced the *no prediction* mode in iCoRA in order to familiarize the users with the tool before being exposed to performance predictions. Furthermore, this condition provided us with a control scenario that allowed us to contrast the effect of the mere presence of performance predictions on the users, regardless of the chosen visual representation. This course selection task was followed by a questionnaire on the rationale behind the participants' decisions.

Four course selection tasks with prediction. The *no prediction* course selection task was followed by four others, each of which presented the predicted grades through the visual representations of a single type (*specific* or *vague*). Due to our within-subject study for the representation type, each participant was exposed either to the *specific* visualizations or to the *vague* ones. The association participant-representation type was done randomly, before the execution of the study. We used Latin squares to balance the order in which each participant saw the corresponding representations.

All the course selection tasks were based on the same academic program and the history of the same fictional student. However, the courses available for enrollment differed among tasks. Following the strategy used in Study 1, we introduced courses from other CS curricula at semesters six and seven of the academic program shown to our participants. 9 out of the 11 introduced courses were available for enrollment and were distributed uniformly among the three course categories defined by ESPOL. Each category contained a hard course, one of average difficulty, and an easy one. The introduced courses were unique to a selection task. That is, in every task, participants would see a different set of available courses that had not appeared before and would not appear in subsequent tasks. Under the hood, however, the set of available courses was the same in terms of type, prerequisites, workload, difficulty, historical distribution of grades, and underlying prediction model. Only the
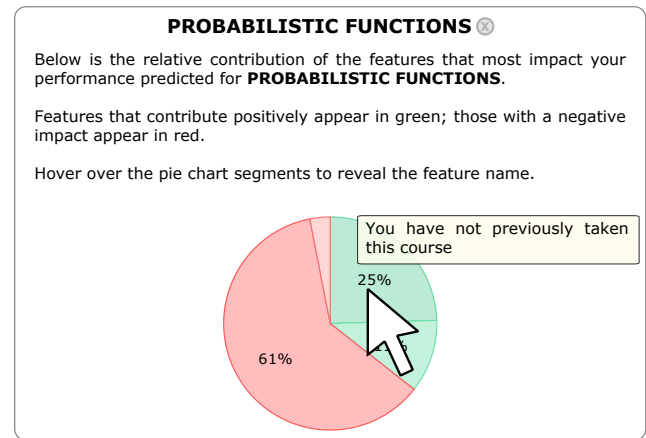

PROBABILISTIC FUNCTIONS ⊗

Below is the relative contribution of the features that most impact your performance predicted for **PROBABILISTIC FUNCTIONS**.

Features that contribute positively appear in green; those with a negative impact appear in red.

Hover over the pie chart segments to reveal the feature name.

You have not previously taken this course

25%

61%

Figure 6: Simplified explanation shown by the *Why* button in the version of iCoRA used in Study 2

names and the position of the courses within the program were different. For example, the course *Micro- & Nanotechnologies* that appeared in the value visual representation had the same features as the *Data Protection* course of the text representation. We made this decision to make the students' chosen sets comparable across different course selection tasks.

Each course selection task was followed by the same rationale questionnaire used after the *no prediction* task.

Closing questionnaire. The experiment concluded with a final questionnaire asking participants about their preferences on how iCoRA presented its predictions.

Based on the observations and participants' comments of Study 1, for this study we simplified the explanations shown by the *Why* button. Specifically, we removed the textual summary. Figure 6 depicts how iCoRA's prediction explanations looked like in this study. We also hid from the interface the section that shows the changes in the student's GPA (Figure 1g) in order to study the influence of the predicted grades in isolation.

6.5 Data Collection and Statistical Tests

Besides the questionnaires answers, we recorded the set of courses our participants chose in each selection task, as well as their associated grades and workload. Because this study did not involve interviews or screen recordings, we instrumented iCoRA to log the consequences of several types of user interactions. These included the partial sets participants progressively built when deciding on their courses, as well as the number of times they opened the explanations of the predicted performances through the *Why* button and the duration of these events.

To inquire whether the prediction representations explain the differences in the means of the dependent variables mentioned above (students' decisions and behavior), we carried out two types of statistical tests. We conducted a **within-subject analysis** with the data of the students exposed to each experimental condition—*specific* or *vague*. For these analyses we used one-way ANOVAs, per

dimension of each dependent variable. When the data was found to be not spherical (i.e., the Mauchly's test failed), we applied a Greenhouse-Geisser correction. All post-hoc tests were corrected for multiple comparisons using Bonferroni corrections. These analyses also included the non-predictive measures obtained under the *no prediction* condition, since all participants were exposed to it.

The **between-group analysis** was carried out using a series of t-tests. These analyses consisted of a cross comparison between the measurements of the dependent variables under each representation of the experimental conditions. This yielded a set of 80 comparisons (e.g., average number of selected courses using: value and text, value and area, value and color, value and faces, and so on).

We also carried out a **contrast itemset mining analysis** [42, 55] to investigate whether some prediction representations may have induced students to select particular groups of courses.

All of our tests were carried out with a significance level $p < 0.05$.

6.6 Results

We excluded the data of 12 participants from our analyses due to inconsistencies between their answers and the usual enrollment habits of ESPOL students[2]. Our analyses are thus based on the data of 79 participants—37 who were exposed to the *vague* representations and 42 who used iCoRA under the *specific* ones.

We present our findings along three axes, namely the students' decisions, their behaviors, and their preferences.

---

[2]These 12 participants chose either more than 6 courses or less than 3. The first scenario is not allowed at ESPOL. On the other hand, enrolling in less than 3 courses mostly happens under very specific circumstances (e.g., when a student is at their very last academic term). Hence, our participants did not have any valid reason to choose so few courses.

6.6.1 Students' Decisions. The decision of a student after a course selection task with iCoRA is defined by the set of courses selected. We elaborate on our findings in three stages. In the first stage, we discuss the results of the within-subject and between-group analyses. In a second stage, we compare the chosen courses with all the partial sets ever composed by the students. This analysis aims at detecting the grade maximization effect observed in Study 1. In a third and final stage, we report the results of an analysis based on contrast itemset mining [42] on the courses chosen by the students. The goal of this analysis is to identify groups of courses that are preferred by the participants exposed to a particular type of visual representation.

Within-subject analysis. Our analysis did not yield any significant differences within the *vague* representations condition for any of the dimensions of our dependent variable. On the contrary, we found significant differences between the **average predicted grade** ($F(3.028, 124.146) = 8.097$, $p = 0.0005$, $\eta_p^2 = 0.165$) within the students exposed to the *specific* representations condition. The post hoc tests revealed significant differences in the means of this dimension between the variant without prediction and each of the *specific* prediction representations (see Table 1a and 1b). Note that the *no prediction* condition reaches a lower mean in the average predicted grade than those in the *specific* representations condition.

We also detected significant differences between the means for the **average chosen workload** ($F(2.905, 119.102) = 5.771$, $p = 0.001$, $\eta_p^2 = 0.123$). The pairwise comparisons revealed differences between the *no prediction* condition and the value and stars representations (see Tables 2a and 2b). Note again, that the mean for the *no prediction* condition is higher than the previously highlighted representations.

	Mean	Std. Deviation
no prediction	7.33	0.39
value	7.72	0.44
range	7.63	0.37
bars	7.62	0.43
stars	7.61	0.35
(a) Means and standard deviations		

	no prediction	value	range	bars	stars
no prediction	1	-.386*	-.298*	-.289*	.279*
value		1	.088	.097	.107
range			1	.009	.019
bars				1	.010
stars					1
(b) Post hoc comparisons. Mean differences shown.					

Table 1: Results of the ANOVA for the average predicted grade of the course sets selected by the students. Post hoc comparisons use Bonferroni adjustment. * shows the mean difference is significant at the .05 level.

	Mean	Std. Deviation
no prediction	4.34	0.57
value	3.84	0.67
range	3.93	0.62
bars	4.02	0.56
stars	3.97	0.51
(a) Means and standard deviations		

	no prediction	value	range	bars	stars
no prediction	1	0.502*	0.413	0.313	0.369*
value		1	-.089	-.188	-.133
range			1	-.099	-.044
bars				1	.056
stars					1
(b) Post hoc comparisons. Mean differences shown.					

Table 2: Results of the ANOVA for the average chosen workload (in hours/week) of the course sets selected by the students. Post hoc comparisons use Bonferroni adjustment. * shows the mean difference is significant at the .05 level.

Between-group analysis. We carried out a set of t-tests between groups of students exposed to both the *specific* and *vague* representations. This round of experiments yielded a significant difference in the **average number of selected courses** for the text (m = 5.225) and bars (m = 4.38) representations, $t(80)$ = 2.09, p = 0.04.

Grade Maximization Effect. As done for Study 1 (Section 5.3.1), we looked at the maximal grades of all the course sets ever composed by a student during an interaction with iCoRA. We then calculated, for each visual representation, the average percentile of the maximal grade of the selected course. The average percentile ranges from 53.37 (standard deviation σ = 29.38) for the stars representation, to 66.44 (σ = 24.30) for the *no prediction* mode. The average percentile across all interactions is 59.42 (σ = 26.64). The trends are similar for the average grade of the course sets, whereas for the minimal grade the highest average percentile is 32.78 (σ = 22.70). These results contest the grade maximization effects we observed for the GPA in Study 1.

Itemset Mining on Courses. We also investigated whether some prediction representations may have leaned students towards choosing specific courses. For this purpose, we looked at the co-occurrence graphs of the courses the students selected per visual representation (Figure 7). The nodes of these graphs represent the courses available for enrollment in the study's selection tasks. Thicker edges denote higher co-occurrence. We observe some recurrent cliques in all scenarios, e.g., the set { 1, 2, 4, 6 } is prominent in almost all cases. Motivated by this insight, we carried out a deeper analysis based on contrast itemset mining [42, 55]. This technique finds groups of courses that co-occur more frequently in a visual representation than in others. We measure the relevance of those groups via the *growth ratio* score [42], which given two categories, defines the ratio of the frequencies[3] of a group of courses in each of the two categories. Values larger than 1 denote "interesting" groups.

---

[3]That is, the number of course sets that contain the group divided by the total number of course sets in the category. We considered groups of courses occurring in at least 10 course sets.
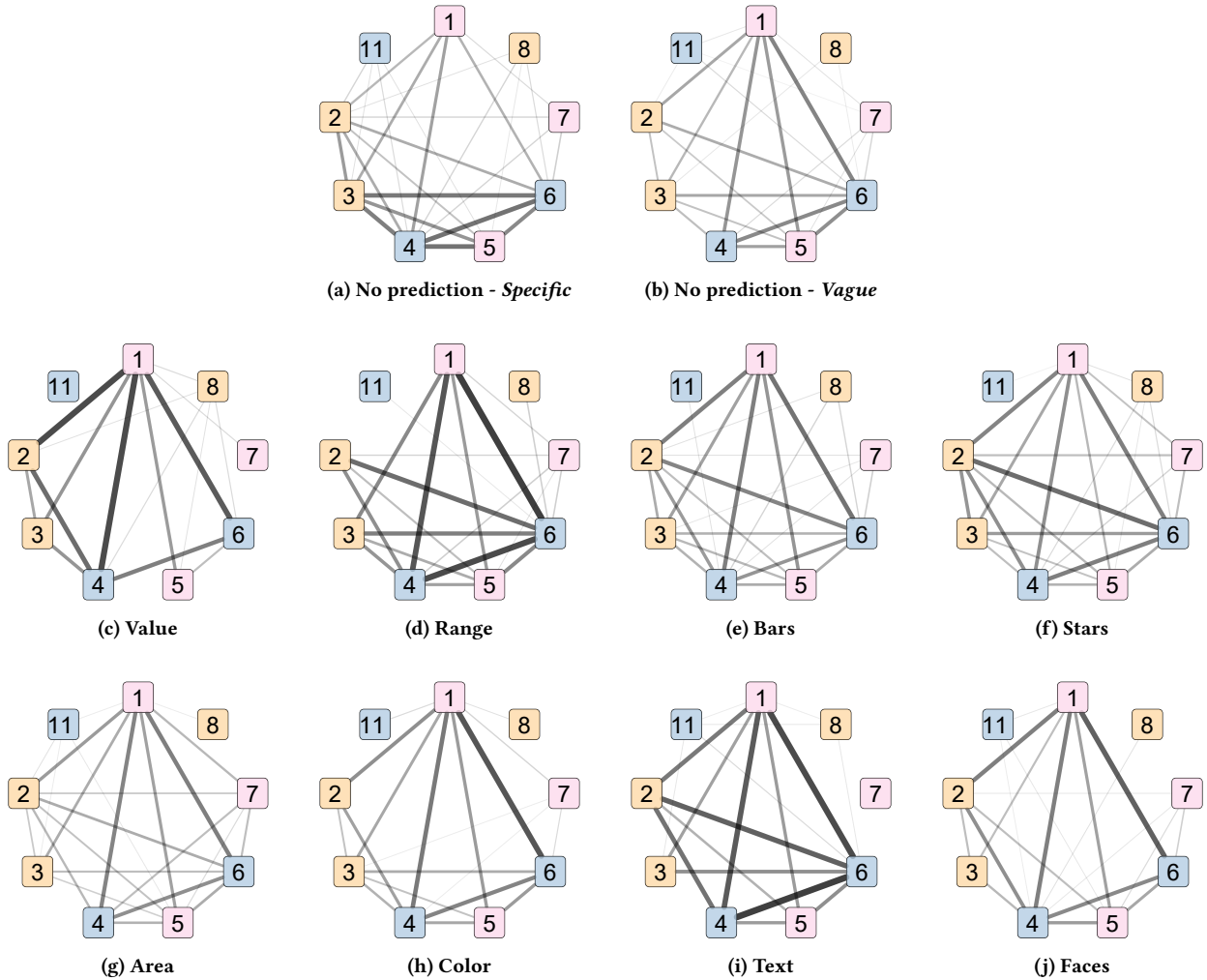

(a) No prediction - *Specific*

(b) No prediction - *Vague*

(c) Value

(d) Range

(e) Bars

(f) Stars

(g) Area

(h) Color

(i) Text

(j) Faces

Figure 7: Co-occurrence graphs of the courses students selected when exposed to different types of prediction representations.

Albeit frequent everywhere, the course group { 1, 2, 4, 6 } is 4.73 times more frequent—with 95% confidence interval (CI) (2.85, 7.81)—in the *no prediction* scenario than in the scenario with the text representation. Similar scores can be found between the scenario with *no prediction* and the faces visual representation—growth ratio 3.68 with CI (2.15, 6.34). The set { 1, 2, 5, 6 } is prominent in the scenario without prediction as it is 5.25 times more frequent—CI (1.78, 15.46)—than for the text representation. Conversely the group of courses { 3, 4, 5, 6 } is 2.86 times more frequent—with CI (1.67, 4.92)—in the bars and faces representations than in the scenario with *no prediction*. The growth ratio scores were calculated on sets of courses selected by disjoint groups of students.

6.6.2 Process. We present the results in regards to this variable in line with our within-subject and between-group analysis protocols. We did not find significant differences for the dimensions of behavior in our within-subject analysis. In regards to the between-group analysis, our t-tests revealed significant differences in **the average number of times the users opened the explanations** for the (i) the text (m=4.78) vs. the value representations (m=2.4, t(51) = 2.44, p = 0.02); (ii) text (m=4.78) vs. bars (m=2, t(53) = 3.04, p = 0.004); (iii) faces (m=4.04) vs. value (m=2.98, t(42) = 2.13, p = 0.04); and (iv) faces (m=4.04) vs. bars (m=2.28, t(46) = 2.94, p = 0.006).

6.6.3 Preferences. We asked students, in the closing questionnaire, their opinion on the most and least appropriate prediction representations. As summarized in Figure 8, there seems to be a consensus when it comes to their preferred prediction representations: 76% of the students favored the range within the *specific* types, whereas 74% of the students deemed the text the most suitable to convey predictions among the *vague* representations.

Preferred Representations. When explaining their preferences, most participants who favored the range representation compared it with the value one and highlighted the capacity of the former to convey uncertainty. This was often mentioned as a booster of their trust in the grades predicted by the tool: "*A range implies that the prediction is subject to uncertainty and it gives a more realistic view than a specific grade. It seems to me a better approach to show the predictions. It seemed more credible and gave me more information than the other options.*" [P40S]; "*I find it difficult to believe that I will get exactly the grade shown by the exact value. However, a range seems more credible to me.*" [P32S]; "*It is better for students to see a range of possible values for their grades, since this indicates how much our grades may vary if we do not keep our effort level; this message is impossible to convey with an exact value.*" [P10S]; "*I consider it more reliable, as it shows a margin in which my grade will be located. I can*

consider that at least I have a margin of error [...] Compared to the other ones, despite being very visual, they don't tell me much at the end of the day.*" [P05S]; "*It is better to have a range, I can't trust an exact grade.*" [P24S]; "*It gives us more confidence because we know that a prediction has a margin of error.*" [P37S]; "*I feel that with an exact grade there is more possibility of error. Instead, a range lets me know, more or less, the grade I may get.*" [P02S].

The students who used iCoRA with the *vague* representations preferred the text mainly because of its simplicity and directness: "*It is much clearer and more explanatory.*" [P03V]; "*It is simple and concise.*" [P09V]; "*I prefer to be presented with things in a more direct way, and this message is, to some extent, encouraging.*" [P10V]; "*It is faster to grasp.*" [P17V]. Students also highlighted that, compared to the others from the *vague* category, the text representation does not require interpretation: "*It is easier to understand, it leaves nothing to interpretation.*" [P45V]; "*It gives me an answer that is easy to understand. With the colors or the faces, I have to infer what the symbols are and what each means.*" [P11V]; "*It seems the most appropriate to me because no previous explanation is needed to understand how it works. It is intuitive and it tells directly how I would perform in a course.*" [P30V]. Other students commented on how close the textual messages were to the advice provided by their advisors or other students: "*I felt that in a certain way, it encouraged me to take the courses, because the language used is similar to the one a friend from my degree would have used when talking about the courses.*" [P28V]; "*The information is somewhat similar to the recommendations my advisor would give me in person.*" [P42V].

Non-preferred Representations. The opinions about the least appropriate representations were varied. The stars and faces stand out as the least preferred representations according to 36% and 39% of the participants. They are followed by the area and the bars, both appearing in 24% of the answers.

In the *specific* category, the stars and the bars were deemed as not precise enough, distracting, and even "not serious" for an academic context. The comments for these representations also highlighted the lack of a numeric representation of the predicted grades, as illustrated by this exemplary statement: "*It does not provide much feedback. What I want to see is my grade.*" [P02V].

The faces were rejected because it was hard for students to discern differences between the representations shown. It was common for students to state that it was hard to distinguish the degree of happiness or sadness in the facial expressions. A similar problem was reported about the circular marks used in the area representation. The sizes of the circles were considered not easily distinguishable.
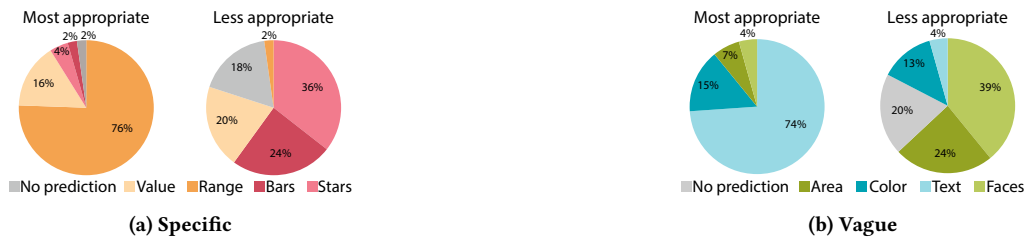

(a) Specific

(b) Vague

Figure 8: Most and least appropriate prediction representations according to the participants' opinions.

7 DISCUSSION

Our discussion is initially structured along the three same axes used to present our experimental results in light of our research questions **RQ1** and **RQ2** (Sections 7.1–7.3). Additionally, in Section 7.4, we discuss the ethical considerations of using GPA-based predictions for course selection and recommendation.

7.1 Students' Decisions

Study 1 tackles **RQ1** by investigating the effect on students of displaying performance predictions during term planning. The observations of this study suggest that predictions can make students embrace a grade and GPA optimization approach, disregarding other important factors (Section 5.3.1). Study 2 inquired whether this behavior was caused by the mere presence of individual course performance predictions, and whether those predictions induce framing effects on the students (**RQ2**). The results of this second study (Section 6.6.1) suggest that, at least for the individual grades, predictions per se do not induce a grade maximization effect. We remark, however, that Study 2 left out the prediction of the GPA. This raises the question of whether this factor might have been the trigger of the maximization effect observed in Study 1, as it has a greater impact on the career of the students than the individual course grades of an academic term.

Study 2 also suggests that the students' decision process and their final choices are indeed influenced by the type of prediction representation (**RQ2**). While no visual representation seems to have favored a grade maximization effect, the *specific* representations seemed to have leaned students towards more optimistic predictions and lighter workloads (see Tables 1b and 2b). This was not the case for the visualizations located at the *vague* end of our spectrum of prediction representations. All this suggests that when exposed to "countable" predictions, students put more effort on the course selection task. This is confirmed by the fact that, on average, the students composed more partial sets when exposed to the *specific* representations: The value and range representations lead the way with the longest sequence of interactions—8.02 and 7.73 sets on average—before making a decision (the total average is 6.59). This indicates that *specific* (i.e., countable) visual representations make students iterate more over their enrollment options, which could be a sign of a deeper and more critical reflection process.

If the courses chosen by the students do not point to a grade maximization effect, then the students must be also taking the workload into account. This assertion is suggested by our itemset mining analysis. The group of courses { 1, 2, 4, 6 }, which is prevalent in all the levels of our experimental conditions, includes 3 courses with a workload of 3 hours each, and one course with a workload of 5 hours. This "formula" actually corresponds to the lightest possible combination of courses in terms of workload. Nonetheless, this logic does not apply to all popular groups of courses. For instance, the itemset { 3, 4, 5, 6 }—prevalent mostly in the *vague* representations bars and faces—leads to a high workload. Conversely, the popularity of this set can be explained by the location of its components in the visualization of the academic program: These appeared together at the left-most end of the upper row of courses available for enrollment. A similar observation applies to the group of contiguous courses { 1, 2, 5, 6 }, particularly prominent in the

no prediction scenario. Altogether, this evidence suggests that the students do care about workload when no prediction about their GPA is shown, and that the *vague* representations may induce students to think less over their enrollment choices—an overreliance effect. Furthermore, we highlight a preference for courses in the upper row of available courses in the program. This conforms to the strategy expressed by some students in Study 1 and also confirmed by the demographics questionnaire of Study 2: when deciding on their enrollment, students favor courses located at the level of the upcoming semester in their study program.

7.2 Course Selection Process

We discuss this axis in terms of two aspects of the course selection process, namely, the strategy used by the students and their interactions with iCoRA's explanations. We also discuss the potential risk of overreliance and automation-complacency effects.

Strategy. Our analysis of the students' interactions with iCoRA showed a recurrent two-stage strategy to compose set of courses. In the first stage, students generally added three courses to the grades prediction panel. This was followed by an exploration phase in which they added and removed courses repeatedly. This behavior was common regardless of the prediction visual representation (even in the *no prediction* scenario). While this might imply students picked up and removed courses driven by some sort of optimization objective, our analysis of Section 6.6.1 indicates that they did not necessarily settle for the most optimistic predictions. The lower average workloads observed for the *specific* visual representations reiterates the role of the workload in the students' approach.

Interactions with the Explanations. In Study 1, our participants exhibited a weak interest on the explanations iCoRA provided, both for the course difficulty estimators and the performance predictions (accessible via the *Why* button). Nevertheless, we obtained hints about possible causes of that lack of interest. Some of the students argued that the explanations were too long and far from obvious, which is consistent with the observation that the first interaction was comparatively long (53 seconds on average) and was rarely followed by a second one. When designing our second study, we took action in this regard and simplified the performance explanations. This, however, did not increase the interest of the students: In 49% of the course selection tasks, our participants did not interact with the performance explanations at all, and only in 11% of the cases there was more than one interaction. The total time invested in reading the explanations was on average 14 seconds, although we expect it to be shorter than for Study 1 since the explanations were more concise. The within-study and between-group analyses described in Section 6.6.2 did not show any significant difference in the total interaction time with the explanations as a consequence of the visual representation of the prediction. However, the *vague* representations color and faces led to significantly more interactions than the *specific* visualizations value and bars. These higher numbers of interactions with iCoRA's explanations could be explained by the interpretation overhead incurred by the *vague* representations. Nevertheless, explanations might have been less needed for the *specific* representations, as these make the grades predicted more directly readable.

Overreliance and Potential Complacency Effects. The lack of interest in the explanations provided by iCoRA might suggest some sort of automation complacency [29] in the students regarding iCoRA's performance predictions. Although it could be said that students did not need major explanations—because they trusted the system—our analyses rather suggest that they were, in most of the cases, not very interested in understanding what was happening inside the tool. Explanations incur, however, a cognitive load on users. A promising research direction could be to decide the right stages of the course selection process where explanations are pertinent and desirable. An exciting venue for future research would be the exploration of student-generated explanations of their performance. This type of explanation design has shown benefits in the visualization of complex scientific phenomena [60].

The evidence we gathered on overreliance and potential complacency effects, however, is not conclusive. Further investigations are needed in this area to better understand and fully characterize these effects.

7.3 Preferences

The preferences of the students elicited via our studies indicate that they value two attributes in performance predictions, namely *credibility* and *directness*. The first factor is corroborated by their preference for the range (overall, the most preferred *specific* representation) over the value. Indeed, students rated predictions with ranges as more reliable and credible than exact values, as ranges convey more information. These observations are consistent with existing studies of the cognitive preferences of people regarding AI agents [27]. The evidence suggests that, from the perspective of user acceptance, the plausibility of a prediction or explanation, i.e., its concordance with the users' background and common sense, is as important as its comprehensibility or simplicity. This credibility dimension may also explain the preference of students for the text over the other *vague* representations, although this preference can also be explained by the directness of textual predictions. This was explicitly stated by several participants of Study 2, who valued textual predictions as direct, simple, and easy to understand. However, it is equally plausible that textual recommendations generated more trust because they expressed messages that were close to what a human advisor or colleague would say.

7.4 Ethical Considerations of Grade- and GPA-based Predictions

An important ethical concern of our work arises from profiling students based on their course grades and GPA. Such predictions have been discussed as a potential threat to the students' potential and self-efficacy [9, 31] and must be tempered with caution. As we mentioned earlier, iCoRA does not seek to replace the human advisors. Rather, it intends to support and facilitate the student-advisor dialogue through a data-driven approach. There exist factors outside the student's academic environment that performance-based predictive models cannot account for (e.g., extracurricular activities, family and health issues). Thus, we highlight the need for human judgment on top of *any* data-based academic performance prediction. Our observations suggest that the decisions and recommendations derived from iCoRA—and similar tools—must remain

on the human side of the academic advising process. This is a key aspect of interactive visualization technologies where "humans in the loop" make decisions and perform analytical tasks based on data. That being said, it is also important to highlight that even with human intervention, overreliance effects may arise. Moreover, students ultimately decide their enrollment based on factors that may change after their advising meetings (e.g., availability of places in courses, scheduling constraints). These aspects are beyond the advisors' reach and are often handled exclusively by the students, at the exact moment of enrollment. Thus, more than recommending which courses could be taken, advisors should provide students with guidelines on the criteria to consider when deciding on their enrollment. Performance and workload should not be the only factors to observe, and other considerations will likely include the specific context of each HEI and the personal circumstances of each student.

We also highlight the reductionist nature of the GPA in describing the students' performance. After all, the learning process comprises other aspects—and, hence, other metrics (e.g., development of learning outcomes, levels of engagement, the students' learning style, or teaching preferences)—that might be more suitable for prediction. Such metrics, however, are rarely systematically collected by HEIs. When measured, they are often kept by instructors to reflect on specific, localized activities. Therefore, they seldom become part of the students' official academic record. At ESPOL, for example, the GPA is the official performance metric that students are exposed to throughout their career. It was also the only performance indicator readily available for prediction. This practical limitation forced us to study iCoRA with a focus on GPA.

One promising future perspective of this work is to elicit conversations with policymakers on alternative performance metrics that could be gathered at HEIs to further empower students and counselors in their use of educational data. Learning strategies are increasingly more oriented to emphasize learning through the demonstration of what a student is able to do with the knowledge they acquire or develop [35]. Moreover, several studies suggest that how students are assessed impacts their learning performance [1, 54, 62]. Therefore, assessment should also be focused on measuring the learning quality, rather than the learning quantity [64]. To stress these aspects, alternative assessment activities require students to demonstrate of thinking and problem-solving skills, involvement or engagement, performing a significant task, creating an artifact or product, etc. They also resort to portfolios, case-based or peer assessments, and observation of students group process [46, 64].

Given this variety of assessment activities, there is a myriad of indicators that could be used as proxies of student performance e.g., level of engagement in meaningful activities, quality of the interactions between peers in collaborative tasks, reflections about students' learning during a design/creation process, certificate or badge achievements, student outcomes observation when doing an activity or working in groups. The increasing penetration of LMS, MOOCs, and learning apps may enable monitoring these indicators and include them as part of the data students and advisors could visualize and discuss during their meetings. However, until enough data on alternative metrics is available, our findings should be interpreted considering the limitations of the GPA discussed throughout the paper.

8 LIMITATIONS AND OPEN QUESTIONS

An obvious threat to the ecological validity of our studies is the use of a fictional academic history and courses from external CS curricula. We acknowledge that the stakes are higher in real-world scenarios, where poor enrollment decisions have a real impact on the students' life. That being said, we did not find any indication that introducing courses from other curricula in our studies got in the way of our participants' decisions. Although it could be argued that these decisions might have been made without much consideration, our video analysis of the data from Study 1 showed that students indeed engaged in the course selection tasks, often thinking aloud about their enrollment options. The setting of Study 2 did not allow for the collection of video data. However, our quantitative analyses were based on the data of participants whose decisions complied with the usual enrollment patterns of ESPOL students. The use of synthetic data also allowed us to reduce the *vox populi* effect, toward a more objective course selection process.

It is important to remark that course selection is also affected by non-academic aspects. Extracurricular workload, health and family issues, and many other factors play a role in the decisions and academic performance of students [36]. In this regard, our results should be taken with a grain of salt. In the same vein, the evidence gathered through our studies is not sufficient to rule out the presence of extra-representational factors (e.g., preferences, conventions) that can also influence the interpretation of a visual representation [34] and, thus, the decisions of the viewer.

Our findings may also be limited by the background of our participants. CS students are familiarized with visual representations of data and scientific concepts. Additional studies are needed to explore whether the effects we observed hold for students from different backgrounds.

Moreover, our observations of the grades maximization effect are not conclusive. Given the evidence we gathered, our intuition is that, in the presence of predictions for the GPA, students tend to look for maximization, but not when only individual course grades are shown. This question, however, is yet to be answered. Additional studies (e.g., in-situ pilots) are also needed to issue more concrete design recommendations for future course selection tools. iCoRA is a high-fidelity prototype, but its institutional deployment is still subject to the outcomes of multiple studies and discussions with several parties. The use of learning analytics dashboards should be thoroughly analyzed before their adoption at HEIs.

Finally, our findings highlight that the design of a tool like iCoRA must consider the role that both students and advisors play in the course selection process. In line with the goal of interactive visualization technologies, iCoRA and similar tools require humans in the loop and this requirement should not be underestimated. Otherwise, this type of technologies run the risk of being perceived as oracles that people are supposed to trust and never question.

9 CONCLUSION

This paper investigated the effects of performance predictions on students when they plan their upcoming term. To this end, we used iCoRA, an interactive visualization tool that enables the composition of arbitrary sets of courses and provides performance predictions and explanations.

A qualitative study of the tool found that in response to performance predictions for both individual course grades and the GPA, students tend to approach course selection as a performance maximization problem, even to the detriment of other factors such as the workload. We also observed little interest in understanding the rationale behind the predictions provided by the tool.

In a follow-up quantitative study, we investigated whether the maximization and overreliance effects were affected by the type of visual representation used to convey iCoRA's performance predictions. To this end, we designed a *specific* to *vague* spectrum of visual representations for performance predictions. In this second study, we did not found evidence of maximization effects when the GPA is not shown together with the individual course grade predictions. The participants' lack of interest in the explanations, however, persisted. We also found several significant differences in aspects such as the average predicted grade and workload of the selected courses. These differences arose both among visual representations of the same type and between different types.

Our observations show that framing effects arise when visual structures are used to communicate performance predictions to students. That is, some of the visual representations we studied have the potential to shape the students' decisions and their decision process. Furthermore, specific types of visual representations elicit strong preferences and aversions on the students. These observations are of great value to design better data-driven course selection tools. Equally importantly, our insights provide new empirical evidence on how different design choices can shape the way people interpret visual representations of data.

REFERENCES

[1] The Quality Assurance Agency. 2001. *QAA Code of Practice for the assurance of academic quality and standards in higher education: Career Education, Information and Guidance (CEIG).* Taylor & Francis. https://doi.org/10.11120/plan.2001.00020026

[2] Stephen J. Aguilar, Stuart A. Karabenick, Stephanie D. Teasley, and Clare Baek. 2021. Associations between Learning Analytics Dashboard Exposure and Motivation and Self-regulated Learning. *Computers & Education* 162 (2021), 104085. https://doi.org/10.1016/j.compedu.2020.104085

[3] Thushari Atapattu, Katrina Falkner, and Hamid Tarmazdi. 2016. Topic-Wise Classification of MOOC Discussions: A Visual Analytics Approach. In *Proceedings of the 9th International Conference on Educational Data Mining.* International Educational Data Mining Society.

[4] Albert Bandura, M Rumsey, C Walker, and J Harris. 1994. Regulative function of perceived self-efficacy. *Personnel selection and classification* (1994), 261–271.

[5] Talia Bar, Vrinda Kadiyali, and Asaf Zussman. 2008. Quest for Knowledge and Pursuit of Grades: Information, Course Selection, and Grade Inflation. *Behavioral & Experimental Economics* (2008). https://doi.org/10.2139/ssrn.1019580

[6] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful Junk?: The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10).* ACM, New York, NY, USA, 2573–2582. https://doi.org/10.1145/1753326.1753716

[7] Eric P. S. Baumer, Francesca Polletta, Nicole Pierski, and Geri K. Gay. 2017. A simple intervention to reduce framing effects in perceptions of global climate change. *Environmental Communication* 11, 3 (2017), 289–310. https://doi.org/10.1080/17524032.2015.1084015

[8] Eric P. S. Baumer, Jaime Snyder, and Geri K. Gay. 2018. Interpretive Impacts of Text Visualization: Mitigating Political Framing Effects. *ACM Transactions on Computer-Human Interaction* 25, 4, Article 20 (Aug. 2018), 26 pages. https://doi.org/10.1145/3214353

[9] Scott Beattie, Carolyn Woodley, and Kay Souter. 2014. Creepy analytics and learner data rights. *Rhetoric and reality: Critical perspectives on educational technology. Proceedings ascilite* (2014), 421–425. http://www.academia.edu/download/35842981/69-Beattie.pdf

[10] Sanam Shirazi Beheshitha, Marek Hatala, Dragan Gašević, and Srećko Joksimović. 2016. The Role of Achievement Goal Orientations When Studying Effect of Learning Analytics Visualizations. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (Edinburgh, United Kingdom) *(LAK '16)*. Association for Computing Machinery, New York, NY, USA, 54–63. https://doi.org/10.1145/2883851.2883904

[11] Robert Bodily and Katrien Verbert. 2017. Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. *IEEE Transactions on Learning Technologies* 10, 4 (2017), 405–418. https://doi.org/10.1109/TLT.2017.2740172

[12] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico Bertini. 2017. Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 5462–5474. https://doi.org/10.1145/3025453.3025512

[13] Richard E Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage.

[14] Susan Bull, Michael Kickmeier-Rust, Ravi K. Vatrapu, Matthew D. Johnson, Klaus Hammermueller, William Byrne, Luis Hernandez-Munoz, Fabrizio Giorgini, and Gerhilde Meissl-Egghart. 2013. Learning, Learning Analytics, Activity Visualisation and Open Learner Model: Confusing?. In *Scaling up Learning for Sustained Impact*, Davinia Hernández-Leo, Tobias Ley, Ralf Klamma, and Andreas Harrer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 532–535. https://doi.org/10.1007/978-3-642-40814-4_51

[15] Jaime Castells, Mohammad Poul Doust, Luis Galárraga, Gonzalo Gabriel Méndez, Margarita Ortiz-Rojas, and Alberto Jiménez. 2020. A Student-oriented Tool to Support Course Selection in Academic Counseling Sessions. In *Proceedings of the Workshop on Adoption, Adaptation and Pilots of Learning Analytics in Underrepresented Regions, co-located with the 15th European Conference on Technology Enhanced Learning 2020 (ECTEL 2020)*. http://ceur-ws.org/Vol-2704/paper4.pdf

[16] Jonathan P Caulkins, Patrick D. Larkey, and Jifa Wei. 1996. Adjusting GPA to Reflect Course Difficulty. (1 1996). https://doi.org/10.1184/R1/6470981.v1

[17] Tomas Chamorro-Premuzic and Adrian Furnham. 2008. Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences* 44, 7 (2008), 1596 – 1603. https://doi.org/10.1016/j.paid.2008.01.003

[18] Sven Charleer, Andrew Vande Moere, Joris Klerkx, Katrien Verbert, and Tinne De Laet. 2018. Learning Analytics Dashboards to Support Adviser-Student Dialogue. *IEEE Transactions on Learning Technologies* 11, 3 (2018), 389–399. https://doi.org/10.1109/TLT.2017.2720670

[19] Sorathan Chaturapruek, Thomas S. Dee, Ramesh Johari, René F. Kizilcec, and Mitchell L. Stevens. 2018. How a Data-Driven Course Planning Tool Affects College Students' GPA: Evidence from Two Field Experiments. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (London, United Kingdom) *(L@S '18)*. Association for Computing Machinery, New York, NY, USA, Article 63, 10 pages. https://doi.org/10.1145/3231644.3231668

[20] Amar Cheema and Rajesh Bagchi. 2011. The Effect of Goal Visualization on Goal Pursuit: Implications for Consumers and Managers. *Journal of Marketing* 75, 2 (2011), 109–123. https://doi.org/10.1509/jm.75.2.109 arXiv:https://doi.org/10.1509/jm.75.2.109

[21] Herman Chernoff. 1973. The Use of Faces to Represent Points in k-Dimensional Space Graphically. *J. Amer. Statist. Assoc.* 68, 342 (1973), 361–368. https://doi.org/10.1080/01621459.1973.10482434

[22] Dennis Chong and James N Druckman. 2007. A theory of framing and opinion formation in competitive elite environments. *Journal of communication* 57, 1 (2007), 99–118. https://doi.org/10.1111/j.1460-2466.2006.00331.x

[23] William S Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. https://doi.org/10.2307/2288400 arXiv:arXiv:1011.1669v3

[24] William S. Cleveland and Robert McGill. 1987. Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society. Series A (General)* 150, 3 (1987), pp. 192–229. http://www.jstor.org/stable/2981473

[25] James N Druckman. 2001. The implications of framing effects for citizen competence. *Political behavior* 23, 3 (2001), 225–256. https://doi.org/10.1023/A:1015006907312

[26] Damien S Fleur, Wouter van den Bos, and Bert Bredeweg. 2020. Learning Analytics Dashboard for Motivation and Performance. In *International Conference on Intelligent Tutoring Systems*. Springer, 411–419.

[27] Johannes Fürnkranz, Tomáš Kliegr, and Heiko Paulheim. 2019. On cognitive preferences and the plausibility of rule-based models. *Machine Learning* 109 (2019), 853–898. https://doi.org/10.1007/s10994-019-05856-5

[28] Dragan Gašević, Amal Zouaq, and Robert Janzen. 2013. "Choose Your Classmates, Your GPA Is at Stake!": The Association of Cross-Class Social Ties and Academic Performance. *American Behavioral Scientist* 57, 10 (2013), 1460–1479. https://doi.org/10.1177/0002764213479362

[29] Kate Goddard, Abdul Roudsari, and Jeremy C Wyatt. 2012. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19, 1 (2012), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

[30] Linda Green and Gul Celkan. 2011. Student demographic characteristics and how they relate to student achievement. *Procedia - Social and Behavioral Sciences* 15 (2011), 341 – 345. https://doi.org/10.1016/j.sbspro.2011.03.098

[31] Wolfgang Greller and Hendrik Drachsler. 2012. Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society* 15, 3 (2012), 42–57.

[32] Fritz H Grupe. 2002. An Internet-based expert system for selecting an academic major: www.MyMajors.com. *The Internet and Higher Education* 5, 4 (2002), 333 – 344. https://doi.org/10.1016/S1096-7516(02)00129-X

[33] Francisco Gutiérrez, Karsten Seipp, Xavier Ochoa, Katherine Chiluiza, Tinne De Laet, and Katrien Verbert. 2018. LADA: A Learning Analytics Dashboard for Academic Advising. *Computers in Human Behavior* 107 (2018). https://doi.org/10.1016/j.chb.2018.12.004

[34] Jessica Hullman, Eytan Adar, and Priti Shah. 2011. Benefitting InfoVis with Visual Difficulties. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec 2011), 2213–2222. https://doi.org/10.1109/TVCG.2011.175

[35] David H. Jonassen. 1994. Thinking Technology: Toward a Constructivist Design Model. *Educational Technology* 34, 4 (1994), 34–37. https://www.learntechlib.org/p/171050

[36] Ahmad A. Kardan, Hamid Sadeghi, Saeed Shiry Ghidary, and Mohammad Reza Fani Sani. 2013. Prediction of student course selection in online higher education institutes using neural network. *Computers and Education* 65 (2013), 1 – 11. https://doi.org/10.1016/j.compedu.2013.01.015

[37] Roger Yao Klomegah. 2007. Predictors of Academic Performance of University Students: An Application of the Goal Efficacy Model. *College Student Journal* 41, 2 (2007).

[38] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174012

[39] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300576

[40] Irwin P Levin, Sandra L Schneider, and Gary J Gaeth. 1998. All frames are not created equal: A typology and critical analysis of framing effects. *Organizational behavior and human decision processes* 76, 2 (1998), 149–188. https://doi.org/10.1006/obhd.1998.2804

[41] Irwin P Levin, Sara K Schnittjer, and Shannon L Thee. 1988. Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology* 24, 6 (1988), 520–529. https://doi.org/10.1016/0022-1031(88)90050-9

[42] Haiquan Li, Jinyan Li, Limsoon Wong, Mengling Feng, and Yap Peng Tan. 2005. Relative risk and odds ratio: A data mining perspective. 368–377.

[43] Lisa Lim, Shane Dawson, Srecko Joksimovic, and Dragan Gašević. 2019. Exploring Students' Sensemaking of Learning Analytics Dashboards: Does Frame of Reference Make a Difference?. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (Tempe, AZ, USA) *(LAK19)*. Association for Computing Machinery, New York, NY, USA, 250–259. https://doi.org/10.1145/3303772.3303804

[44] Phillip Long and George Siemens. 2014. Penetrating the fog: analytics in learning and education. *Italian Journal of Educational Technology* 22, 3 (December 2014), 132–137. https://www.learntechlib.org/p/183382

[45] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777. https://doi.org/10.5555/3295222.3295230

[46] Effie Maclellan. 2004. How convincing is alternative assessment for use in higher education? *Assessment & Evaluation in Higher Education* 29, 3 (2004), 311–321. https://doi.org/10.1080/0260293042000188267

[47] Joyce B. Main and Ben Ost. 2014. The Impact of Letter Grades on Student Effort, Course Selection, and Major Choice: A Regression-Discontinuity Analysis. *The Journal of Economic Education* 45, 1 (2014), 1–10. https://doi.org/10.1080/00220485.2014.859953

[48] Todd McElroy and John J Seta. 2003. Framing effects: An analytic–holistic perspective. *Journal of Experimental Social Psychology* 39, 6 (2003), 610–617. https://doi.org/10.1016/S0022-1031(03)00036-2

[49] Gonzalo Gabriel Méndez, Uta Hinrichs, and Miguel A. Nacenta. 2017. Bottom-up vs. Top-down: Trade-Offs in Efficiency, Understanding, Freedom and Creativity with InfoVis Tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 841–852. https://doi.org/10.1145/3025453.3025942

[50] Gonzalo Gabriel Méndez, Miguel A. Nacenta, and Sebastien Vandenheste. 2016. iVoLVER: Interactive Visual Language for Visualization Extraction and Reconstruction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (Santa Clara, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 4073–4085. https://doi.org/10.1145/2858036.2858435

[51] Tamara Munzner. 2014. *Visualization Analysis and Design* (har/psc edition ed.). A K Peters/CRC Press, Boca Raton. http://www.cs.ubc.ca/~tmm/vadbook

[52] Thomas E Nelson and Zoe M Oxley. 1999. Issue framing effects on belief importance and opinion. *The journal of Politics* 61, 4 (1999), 1040–1067. https://doi.org/10.2307/2647553

[53] Ivana Ognjanovic, Dragan Gasevic, and Shane Dawson. 2016. Using institutional data to predict student course selections in higher education. *The Internet and Higher Education* 29 (2016), 49 – 62. https://doi.org/10.1016/j.iheduc.2015.12.002

[54] Deborah M. Oh, Simeon Slovacek, Susan Tucker, and Ann Hafner. 2003. Assessment Outcomes of Pre-service Teachers. *Assessment & Evaluation in Higher Education* 28, 3 (2003), 279–295. https://doi.org/10.1080/0260293032000059630

[55] Hoang Son Pham, Gwendal Virlet, Dominique Lavenier, and Alexandre Termier. 2019. Statistically Significant Discriminative Patterns Searching. *Lecture Notes in Computer Science* (2019), 105–115. https://doi.org/10.1007/978-3-030-27520-4_8

[56] Andrew J Rotherham and Daniel T Willingham. 2010. "21st-century" skills. *American Educator* 17, 1 (2010), 17–20.

[57] Samara Ruiz, Sven Charleer, Maite Urretavizcaya, Joris Klerkx, Isabel Fernández-Castro, and Erik Duval. 2016. Supporting Learning by Considering Emotions: Tracking and Visualization a Case Study. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. 254–263. https://doi.org/10.1145/2883851.2883888

[58] Richard M. Ryan and Edward L. Deci. 2000. Chapter 2 - When rewards compete with nature: The undermining of intrinsic motivation and Self-Regulation. In *Intrinsic and Extrinsic Motivation*, Carol Sansone and Judith M. Harackiewicz (Eds.).

Academic Press, San Diego, 13 – 54. https://doi.org/10.1016/B978-012619070-0/50024-6

[59] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68. https://doi.org/10.1037/0003-066X.55.1.68

[60] Kihyun Ryoo and Marcia C Linn. 2014. Designing guidance for interpreting dynamic visualizations: Generating versus reading explanations. *Journal of Research in Science Teaching* 51, 2 (2014), 147–174. https://doi.org/10.1002/tea.21128

[61] Cassandra Sample. 2018. Implications of Elective Course Selection by Grade Maximization. (2018).

[62] Richard W Schwartz, James E Burgett, Amy V Blue, Michael B Donnelly, and David A Sloan. 1997. Problem-based learning and performance-based testing: Effective alternatives for undergraduate surgical education and assessment of student performance. *Medical Teacher* 19, 1 (1997), 19–23. https://doi.org/10.3109/01421599709019341

[63] Peter Smith. 1995. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18, 2-3 (1995), 277–310. https://doi.org/10.1080/01900699508525011

[64] Katrien Struyven, Filip Dochy, Steven Janssens, Wouter Schelfhou, and Sarah Gielen. 2006. The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation* 32, 3 (2006), 202 – 222. https://doi.org/10.1016/j.stueduc.2006.08.002

[65] Ding Hooi Ting and Christina Kwai Choi Lee. 2012. Understanding students' choice of electives and its implications. *Studies in Higher Education* 37, 3 (2012), 309–325.

[66] Roger Tourangeau, Mick P. Couper, and Frederick Conrad. 2004. Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly* 68, 3 (09 2004), 368–393. https://doi.org/10.1093/poq/nfh035

[67] Camilo Vieira, Paul Parsons, and Vetria Byrd. 2018. Visual Learning Analytics of Educational Data: A Systematic Literature Review and Research Agenda. *Computers & Education* 122 (2018), 119–135. https://doi.org/10.1016/j.compedu.2018.03.018

[68] Jagoda Walny, Samuel Huron, and Sheelagh Carpendale. 2015. An Exploratory Study of Data Sketching for Visual Representation. *Computer Graphics Forum* 34, 3 (2015), 231–240. https://doi.org/10.1111/cgf.12635

[69] Colin Ware. 2012. *Information Visualization* (3rd revised edition edition ed.). Morgan Kaufmann, Waltham, MA.

[70] Qing Zhou and Fang Yu. 2008. Knowledge-Based Major Choosing Decision Making for Remote Students. In *International Conference on Computer Science and Software Engineering*, Vol. 5. 474–478. https://doi.org/10.1109/CSSE.2008.379