Assessing Critical Thinking Components in Romanian Secondary School Textbooks: A Data Mining Approach to the ROTEX Corpus

Mădălina Chitez[1], Liviu P. Dinu[2,3], Marius Micluța-Câmpeanu[4]
Ana-Maria Bucur[4], Roxana Rogobete[1]
[1]West University of Timișoara, Romania
[2]Faculty of Mathematics and Computer Science, [3]HLT Research Center
[4]Interdisciplinary School of Doctoral Studies, University of Bucharest, Romania
{madalina.chitez,roxana.rogobete}@e-uvt.ro, ldinu@fmi.unibuc.ro
marius.micluta-campeanu@unibuc.ro, ana-maria.bucur@drd.unibuc.ro

Abstract

This paper presents a data-driven analysis of Romanian secondary school textbooks through the lens of Bloom's Taxonomy, focusing on the promotion of critical thinking in instructional design. Using the ROTEX corpus, we extract and annotate almost 2 million words of Romanian Language and Literature textbooks (grades 5-8) with Bloom-aligned labels for verbs associated with pedagogical tasks. Our annotation pipeline combines automatic verb extraction, human filtering based on syntactic form and task relevance, and manual assignment of Bloom labels supported by in-text concordance checks. The resulting dataset enables fine-grained analysis of task complexity both across and within textbooks and grade levels. Our findings reveal a general lack of structured cognitive progression across most textbook series. We also propose a multi-dimensional framework combining cognitive-level and linguistic evaluation to assess instructional design quality. This work contributes annotated resources and reproducible methods for NLP-based educational content analysis in low-resource languages.

1 Introduction

Critical thinking is a key competence in education, shaping students' ability to analyze, evaluate, and synthesize information. It refers to cognitive and metacognitive processes that enable individuals to question assumptions, construct arguments, and engage in logical reasoning (Ennis, 1985). These processes include argument evaluation (e.g., identifying sound reasoning and fallacies), metacognition (e.g., self-monitoring of thinking), and epistemic skepticism (e.g., questioning the credibility of sources and claims). However, recent studies indicate that these very capacities may be eroding in the age of generative AI, which has been shown to reduce users' cognitive effort and reliance on reflective thinking (Lee et al., 2025). In the context of education, the extent to which textbooks promote critical thinking has been a major research concern, particularly regarding curriculum effectiveness (Paul and Elder, 2007) and the role of instructional materials in fostering higher-order thinking skills (Facione, 1990). Studies suggest that educational texts should challenge students intellectually while being cognitively accessible, following developmental frameworks such as the zone of proximal development (ZPD) (Vygotsky and Cole, 1978).

Research on textbooks as facilitators of critical thinking has traditionally relied on manual content analysis and qualitative coding (Halpern, 1998; Kuhn, 2005). However, recent advances in natural language processing (NLP) and computational text analysis have enabled large-scale automated evaluation of educational materials. NLP-based methods allow for the detection and quantification of critical thinking components by analyzing linguistic, structural, and argumentative features in textbooks (Allen et al., 2015; Graesser et al., 2011). These features include argumentative density (e.g., presence of claims, counterclaims, and rebuttals), discourse coherence (e.g., logical connections between ideas), and syntactic complexity (e.g., sentence structures that require higher cognitive processing) (Crossley and McNamara, 2016). The availability of NLP-driven readability and complexity assessment tools (see Section 3) varies across languages, depending on the availability of annotated corpora and computational models designed to process textbook content.

Building on this context, our study is guided by the following research questions:

RQ1: To what extent do Romanian language textbooks at the secondary level include tasks that support higher-order cognitive processes, as defined by Bloom's Taxonomy?

RQ2: How are these tasks distributed across grades and chapters, and do they reflect a coherent

Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications, pages 780–793

pedagogical progression?

RQ3: Which textbook series demonstrate the most effective use of instructional tasks for promoting critical thinking, based on verb-level cognitive classification?

These questions aim to bridge linguistic and cognitive evaluation frameworks through a data-driven analysis of instructional content, contributing both methodological tools and empirical insights to the field of educational NLP.

The paper presents the ROTEX corpus analysis and introduces the computational methodology used to extract, classify, and quantify critical thinking elements in school textbooks.[1] We begin by reviewing related work on critical thinking assessment and educational data mining. This is followed by a description of the corpus and the NLP-based methods used to evaluate cognitive complexity. We then present the results of our analysis, focusing on the patterns found in Romanian school textbooks and their implications for curriculum development. The paper concludes with a discussion on the educational relevance of our findings and the potential for integrating AI-driven tools in textbook evaluation and curriculum design.

2 Related Work

Thinking, in its broadest sense, is an active and deliberate process through which individuals make sense of information (Dewey, 2022), ask relevant and purposeful questions (Nosich, 2005), identify what they do not know (ibid.), and revise their beliefs based on new evidence. Dewey (2022) defines reflective thinking as "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" (p. 6). He further characterizes thought as inherently inferential: "the exercise of thought is, in the literal sense of that word, inference; by it one thing carries us over to the idea of, and belief in, another thing. It involves a jump, a leap, a going beyond what is surely known to something else accepted on its warrant" (p. 26). This view frames thinking not as passive reception, but as a generative act of drawing justified conclusions. These foundational processes represent the basis of more specific forms of thinking, such as critical thinking, which adds a layer of evaluative and reflective judgment. For example, questioning assumptions (Brookfield, 2011), evaluating arguments (Halpern, 2013), and recognizing knowledge gaps (Nosich, 2005) are central to critical engagement. Brookfield (2011) further defines critical thinking as the intentional effort to uncover hidden reasoning structures and challenge taken-for-granted beliefs. Kahneman (2011) complements this view by highlighting the dual-process nature of thinking: fast, intuitive cognition and slow, deliberate reasoning, both of which influence how individuals analyze and respond to information. Together, these perspectives suggest that critical thinking is not separate from general thinking but represents its most reflective, analytical, and self-aware form. To apply these dimensions of thinking in instructional design, educators have adopted structured cognitive frameworks, with Bloom's Taxonomy being the most widely used.

2.1 Bloom's Taxonomy and linguistic research

Bloom's Taxonomy is a foundational framework in pedagogy that categorizes cognitive learning objectives into six hierarchical levels: *remember, understand, apply, analyze, evaluate*, and *create* (Bloom et al., 1956; Anderson and Krathwohl, 2001). These levels provide a systematic approach to designing, analyzing, and evaluating educational materials by addressing varying cognitive demands, from basic recall of facts to complex critical thinking and creative tasks. While its original use was in curriculum development, Bloom's Taxonomy has since been widely applied in educational research and, more recently, in corpus-based studies.

In corpus research, Bloom's Taxonomy has played a crucial role in evaluating the complexity of educational texts and tasks. For instance, Oraviţan et al. (2023) used the ROTEX corpus of Romanian language textbooks and applied Bloom's Taxonomy to categorize linguistic features in writing tasks. The authors extracted n-grams and verb patterns to align tasks with taxonomy levels. They found that higher-order levels, such as creation (e.g., write, design), were overrepresented in comparison to mid-level skills like analysis (e.g., compare, analyze). Similarly, Graves (2017) employed Bloom's Taxonomy to examine how writing assignments across university disciplines vary in their cognitive demands, noting the need for balanced progression across the taxonomy levels.

Bloom's Taxonomy has guided the extraction

[1]The code and the annotated data is available here: https://github.com/mcmarius/ro-textbook-parser

and classification of cognitive processes from textual datasets. Tools like Coh-Metrix (Graesser et al., 2011) enable the identification of linguistic markers, such as cohesive devices and argumentation patterns, that correspond to taxonomy levels. Çavdar and Doe (2012) linked writing tasks explicitly with Bloom's Taxonomy to assess critical thinking skills in argumentative assignments, showcasing how cognitive skills can be measured quantitatively in textual corpora.

2.2 Textbook design

A growing consensus among researchers emphasizes the pivotal role of textbook design in cultivating students' critical thinking abilities, with Bloom's Taxonomy serving as a widely endorsed framework for structuring cognitive development in learning materials. The revised taxonomy from Anderson and Krathwohl (2001) explicitly advocates for a hierarchical integration of cognitive processes, ranging from remembering to creating, in the design of instructional materials, highlighting the necessity for a scaffolded progression within and across units. Studies analyzing textbooks across various contexts have consistently found a dominance of lower-order thinking skills (LOTS), raising concerns about insufficient cognitive stimulation. For example, Miyazaki (2024) found that 97.3% of tasks in a widely used Japanese Grade 8 textbook fell into the *remember, understand*, or *apply* categories, despite national curriculum reforms encouraging *analyze*, *evaluate*, and *create* levels. A similar imbalance was reported by Riazi and Mosalanejad (2010) in Iranian high school and pre-university English textbooks, where lower-order tasks were prevalent, although pre-university materials showed a modest improvement in higher-order inclusion. These findings echo those of Mizbani et al. (2023), who found that in Iran's "Vision 2" textbook, high-order thinking activities were lacking across all four language skills, undermining deeper learning opportunities. These studies reinforce the idea that effective textbooks must not only include all levels of Bloom's Taxonomy but must structure tasks progressively within units and increase the proportion of higher-order thinking tasks by grade level. Such recommendations are further supported by global curriculum standards like those in Japan[2], which now explicitly aim to foster "the ability to think, make judgments, and express oneself", outcomes achievable only through textbooks that prioritize higher-order cognition. As Beauchamp and Kennewell (2010) argue, materials that fail to challenge students beyond information recall risk reinforcing surface learning, rather than equipping learners with the reasoning and creativity needed in a complex, unpredictable world.

[2] https://www.mext.go.jp/en/policy/education/overview/index.htm

2.3 Multitasking and critical thinking

While complex, layered tasks are often intended to simulate real-world problem solving, research suggests that combining multiple cognitive demands within a single assignment may, in fact, hinder the development of critical thinking. According to Cognitive Load Theory (Sweller, 1988; Van Merrienboer and Sweller, 2005), the human working memory has limited capacity and overloading it with too many simultaneous instructional demands can lead to superficial engagement rather than deep learning. This is especially problematic when tasks require students to *analyze*, *evaluate*, and *create* under strict, multi-part conditions, prompting a "checklist mindset" rather than genuine intellectual exploration (Torrance, 2007). Paul and Elder (2007) argue that critical thinking flourishes under conditions of conceptual clarity and reflective inquiry, conditions undermined when students are forced to meet narrowly defined sub-goals in one task. Perkins (2008) similarly notes that such instructional designs often lead to "fragile knowledge," where students complete tasks without fully internalizing the concepts involved. While the Revised Bloom's Taxonomy encourages progression toward higher-order thinking (Anderson and Krathwohl, 2001), this does not imply simultaneous execution of all levels in a single prompt. On the contrary, effective critical thinking tasks are often those that isolate and deepen one cognitive demand at a time, especially at the create and evaluate levels, where open-ended exploration is most essential.

3 Method

3.1 Corpus

The analyses are based on the ROLAT subset of the ROTEX corpus (Chitez et al., 2024), the Romanian Corpus of School Textbooks, which comprises Romanian Language and Literature textbooks currently used in secondary schools in Romania. Notably, there is limited continuity within individual publishing house series, as only *ArtKlett* provides a complete set of textbooks for grades 5 through 8

(Table 1).

Textbook	5th grade	6th grade	7th grade	8th grade
ArtKlett	82,249	80,312	95,968	108,918
Booklet	63,416	93,031	-	-
Corint	58,719	85,836	-	93,214
Litera	55,191	67,895	77,288	-
Intuitext	71,279	78,211	85,143	-
CD Press	47,618	51,801	71,273	-
EDP	-	57,096	64,126	-
Paralela 45	-	88,567	99,002	-
Ars Libri	-	-	62,527	-
Aramis	-	-	75,734	66,294
Total	1,880,708			

Table 1: ROTEX sub-corpus size (no. words)

3.2 Annotation

Due to the lack of resources with verbs annotated using Bloom's Taxonomy labels in Romanian, we have annotated the ROTEX corpus with these labels through a multi-step process. The main computational analysis steps are: (1) task extraction, (2) verb extraction, (3) syntactic filtering, and (4) Bloom-level labeling. The processing pipeline is presented in Appendix A, Figure 4).

First, all tasks were extracted from the ROTEX corpus based on two methods: regular expression heuristics and multimodal prompting with Gemini (Team et al., 2023) (the prompt used is detailed in Appendix A, Table 4). Then, tasks found by both methods were deduplicated. Next, verbs were extracted using the spaCy[3] POS tagger for Romanian. To target pedagogical intent, verbs were filtered to retain only those in second person singular or plural, typically indicative of task instructions (e.g., *scrieți*, *gândiți*, *comparați*). This filtering was performed either automatically with spaCy morphological features or manually by human reviewers. For the latter, a group of trained students identified task-related verbs by examining in-text concordances.

Next, the remaining verbs were reviewed for accuracy by expert annotators on the research team. Bloom's Taxonomy labels were assigned based on a seed list of verbs from existing literature, directly translating verbs from the Bloom Taxonomy Levels (e.g., "analyze" translated to "analizați") or from didactic expertise. Verbs without automatic label matches were manually annotated by the same group of human raters. In-text concordances were again used to support disambiguation, particularly for verbs potentially associated with multiple Bloom categories. A final expert verification ensured the accuracy of the annotations, resulting in the finalized list of verbs and their corresponding Bloom labels. Table 2 shows the most frequent verbs, identified as recurring across all instructional prompts and tasks within the ROTEX corpus.

A multi-step human validation process was implemented to ensure the reliability of the Bloom-level annotations. After the initial annotation phase by trained student raters, the assigned labels were reviewed by expert members of the research team, who verified the alignment between each verb's usage in context and its cognitive category. Special attention was given to polysemous verbs and those that could potentially map to multiple Bloom levels. In these cases, in-text concordances were used to assess task intent and clarify ambiguities. Although no formal inter-annotator agreement metric was calculated, the annotation process was iterative and consensus-based, ensuring consistency in labeling and fidelity to both linguistic form and pedagogical function. It is important to note that assigning cognitive categories to verbs in isolation would be reductive; instead, the full task must be taken into account. Therefore, when applying Bloom's Taxonomy to real educational materials (such as ROTEX), a contextualized approach has to be used, to ensure that the categorization remains pedagogically meaningful and accurately reflects how the tasks are designed to engage students cognitively.

In total, 434 verbs were annotated: 86 were labeled as "*analyze*," 84 as "*apply*," 47 as "*create*," 112 as "*understand*," 26 as "*remember*," and 79 as "*evaluate*". The annotated list of verbs was then compiled at both the chapter and textbook levels to enable macro- and micro-level distribution analysis. The annotated dataset (Bloom-labeled verbs per task) is publicly available via GitHub[4], under a CC-BY license.

[3] https://spacy.io/

[4] https://github.com/mcmarius/ro-textbook-parser

Bloom Taxonomy Level	Task signal phrases
Remember: Recalling facts, concepts, or basic information.	*numește, rememorează, reproduceți, rostește, urmărește* (EN: name, recall, reproduce, recite, follow)
Understand: Explaining ideas or concepts.	*asociază, caută, centralizează, delimitează, descrie, extrage, indică, identifică, menționează, organizează, precizează, recunoaște, selectează, subliniază* (EN: associate, search, cluster, delimite, describe, extract, indicate, identify, mention, organize, specify, recognize, select, underline)
Apply: Explaining ideas or concepts.	*adaugă, adresează, alcătuiește, aplică, arată, combină, completează, construiește, demonstrează, exemplifică, folosește, formează, formulează, îmbină, înlocuiește, încadrează, marchează, rezolvă, transformă, valorifică* (EN: add, address, compose, apply, show, combine, complete/fill in, build, demonstrate, exemplify, use, form, formulate, merge, replace, frame, label, solve, transform, utilize)
Analyze: Identifying connections between ideas or breaking down a concept.	*analizează, aseamănă, caracterizează, comentează, corectează, corelează, definește, desprinde, separă, stabilește, lucrați pe echipe* (EN: analyze, compare, describe/characterize, comment/interpret, revise, correlate, define, distinguish, differentiate, determine, collaborate)
Evaluate: Forming judgments or justifying a decision or opinion.	*alege, argumentează, compară, convinge, dezvoltă, discută, documentează, evaluează, interpretează, justifică, motivează, susține, verifică* (EN: choose, argue, compare, convince, develop, discuss, research, evaluate, interpret, justify, motivate, support, check)
Create: Producing something new or original.	*compune, concepe, confecționează, continuă, desenează, evocă, gândește, imaginează(-ți), închipuie(-ți), prezintă, realizează, redactează, reformulează, rescrie, scrie, transpune* (EN: compose, design, make, complete, draw, evoke, think, imagine, envision, present, create, write, rephrase, rewrite, write, adapt)

Table 2: Bloom-taxonomy verb examples

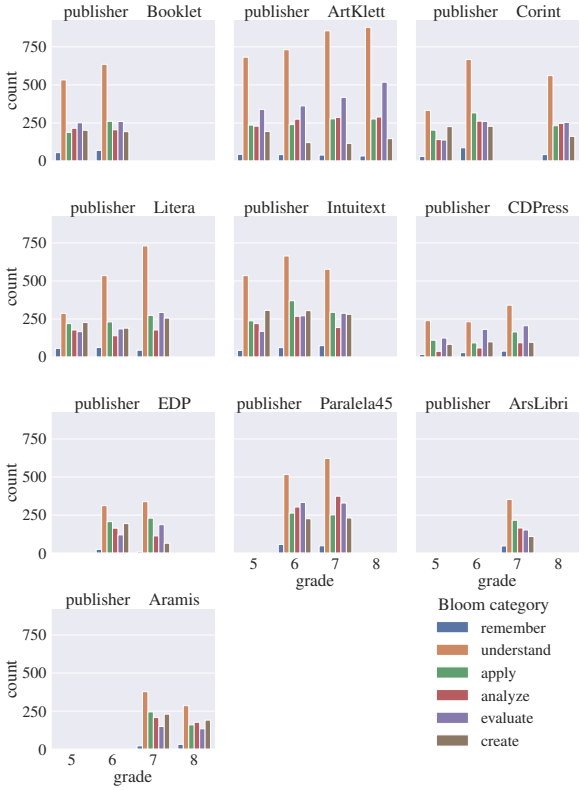

Figure 1: Bloom category counts by publisher and grade

4 Textbooks Analysis

In this section, we present the findings from our analysis of the ROTEX corpus concerning Romanian language textbooks for grades 5 to 8. The analysis indicates a general absence of structured cognitive progression among most publishers. Below, we provide the identified patterns of Task Complexity Distribution (TCD).

4.1 TCD per textbook series

According to Bloom's Taxonomy, instructional tasks should ideally evolve from foundational cognitive levels, i.e. *remember* and *understand*, toward more complex processes such as *analyze*, *evaluate*, and *create*. However, the distribution of task types across publishers and grades, as shown in Figure 1, suggests minimal scaffolding toward higher-order thinking.

Publishers such as *ArtKlett* exhibit consistently high counts of lower-order tasks (*remember* and *understand*) across all four grades (43.3% or 3304 out of 7626), with surprisingly high rates in higher grades (43.7% or 1805 out of 4132), when the upper-order tasks should prevail. Other textbooks, such as *CD Press* and *EDP*, maintain a near-flat profile across grades, with *understand* tasks mak-

ing up 36% of the content (812 out of 2252), and *create* tasks remaining negligible throughout (12% or 281 out of 2252).[5]

When we look at the growth patterns for higher-order thinking tasks, several trends emerge. *Intuitext* stands out for its clear increase in *analyze* and *evaluate* tasks across grades, and the gradual introduction of *create* tasks by grade 8. *CDPress* and *ArtKlett* also show moderate gains in higher-order categories, especially in the upper grades. *Booklet* and *Paralela45* include a range of task types but lack a clear upward trajectory. In contrast, *Corint*, *Litera*, *ArsLibri*, and *Aramis* remain static, with modest presence of higher-order tasks throughout. *EDP* shows some potential, but lacks full grade coverage.

Based on these findings, a hierarchy of textbooks based on their capacity to promote critical thinking can be established. *Intuitext* ranks highest, showing the most consistent inclusion and progression of *analyze*, *evaluate*, and *create* tasks across grades. *CDPress* follows, with moderate presence of higher-order tasks, especially in upper grades. *ArtKlett* shows some higher-order tasks by grade 8, but lacks consistent instructional sequencing. All other textbooks (*Booklet*, *Paralela45*, *Corint*, *Litera*, *Aramis*, *ArsLibri*) display minimal to no critical thinking tasks, remaining focused on lower-order categories (55% or 10150 out of 18518).

Table 3 shows higher-order tasks included in ROTEX textbooks that exhibit variability and inconsistency regarding Bloom's Taxonomy for several reasons (sequence combining several levels or using multiple verbs and tasks), reflecting the complexity of educational tasks and practices.

4.2 TCD per learning unit within textbooks

By making a fine-grained analysis on the distribution of tasks within each textbook per grade (Appendix A, Figure 5), we can notice similar patterns to the overall task complexity distribution. Among all series, *Intuitext* demonstrates the most coherent and intentional progression of task complexity. Across its chapters, there is a visible build-up from lower-order tasks toward *analyze*, *evaluate*, and even *create*, particularly in the upper grades, reflecting a well-structured approach to competence development. *CDPress* presents a similarly structured pattern, with higher-order tasks becoming more prominent in later chapters, suggesting a

[5]See also Figure 3 for counts shown as percentages (Appendix A).

Task example	**Bloom taxonomy level**
a) *Extrage*, din text, enumerația care se asociază peisajului marin. b) Cum se raportează instanța lirică la ideea de patrie? *Argumentează-ți* răspunsul. (EN: a) *Extract*, from the text, the enumeration associated with the seascape. b) How does the lyric instance relate to the idea of homeland? *Give reasons* for your answer.)	Understand Evaluate Create
Cui *consideri* că îi aparțin cuvintele așezate între liniile de pauză din versul „– Și Dumnezeu cunoaște cum vorba și-o păzește –”? Alege una dintre variantele următoare și *motivează-ți* opțiunea: a) personajului, care jură în fața păsărilor domestice pentru a da greutate cuvintelor sale; b) naratorului, care intervine astfel spre a avertiza cititorul că personajul minte. (EN: Who do you *think* the words between the pause lines in the line "– And God knows how the word is spoken and keeps it –" belong to? Choose one of the following options and *give* your reasons: a) the character, who swears in front of the domestic birds to give weight to his words; b) the narrator, who intervenes to warn the reader that the character is lying.)	Analyze Evaluate Evaluate
Amintește-ți ultima călătorie pe care ai făcut-o. *Formulează* enunțuri care să continue următoarele începuturi... (EN: *Remember* the last trip you took. *Make statements* that continue the following beginnings...)	Remember Create
Recitește textul Fascinații de George Șovu și *notează* în caiet o secvență narativă și una descriptivă, *precizând* ce rol au în cadrul textului. (EN: *Reread* the text Fascinations by George Șovu and *write down* a narrative and a descriptive sequence in your notebook, *specifying* their role in the text.)	Remember Understand Evaluate

Table 3: Examples of exercises that exhibit variability and inconsistency regarding Bloom's taxonomy by combining multiple levels in the same task

lightweight but effective progress model. In contrast, *ArtKlett* includes a wide variety of task types, including higher-order ones, in nearly every chapter from the outset. However, the lack of variation or progression across chapters indicates a dense and static design rather than a pedagogically sequenced one. Other textbook series, such as *Booklet, Litera, Corint, Paralela45*, and *Aramis*, show predominantly flat distributions, with chapters consistently focused on lower-order categories like *remember* and *understand*, and little evidence of an intentional cognitive arc. These findings suggest that while some series embed critical thinking tasks, only a few succeed in distributing them progressively and meaningfully throughout the learning units.

4.3 TCD per learning unit within series across grades

A cross-grade analysis of task distribution reveals distinct patterns in how textbook series support cognitive development over time. *Intuitext* is the only series to exhibit a clear upward trajectory in task complexity, with a gradual increase in *analyze*, *evaluate*, and *create* tasks from grades 5 to 7, aligning well with students' developmental stages. *CDPress* also shows moderate progression, with higher-order tasks becoming more prominent in grades 6 and 7, although coverage is limited. In contrast, *ArtKlett* distributes higher-order tasks relatively evenly across all grades, indicating cognitive density without meaningful scaffolding. The remaining series, *Booklet, Litera, Corint, Paralela45, Aramis*, and *ArsLibri*, maintain static task profiles, dominated by lower-order categories throughout, with minimal evidence of progression. These results suggest that most textbook series do not implement systematic cognitive progression across grade levels, potentially limiting their effectiveness in supporting long-term competence development.

4.4 Multi-task presence

Since tasks can contain multiple sentences and phrases, it is difficult to assign a single label when multiple verbs correspond to several Bloom levels. This complexity persists despite the annotation efforts detailed in Section 3.2, which aimed to assign a single Bloom level to each verb. In the previous sections, we addressed this ambiguity to assign a single Bloom level per task by dividing each task into individual sentences and then further splitting

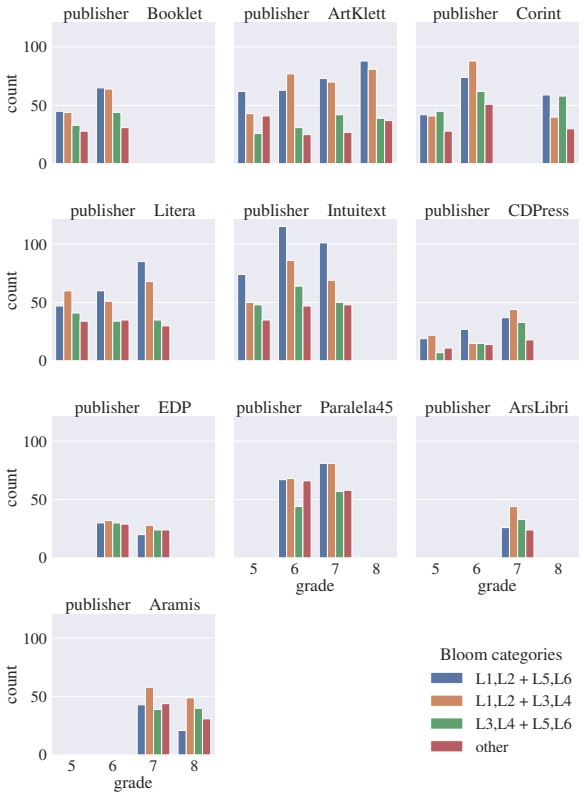

Figure 2: Bloom category counts by publisher and grade for multitask exercises

each sentence based on the main verb identified in the parse tree. Finally, we assigned the label of the first verb in order of appearance.

For this analysis, we keep the labels from all verbs and focus our attention only on these multi-level tasks, which account for 26% of all exercises. After considering tasks at the sentence level, we are left with 15% multi-task exercises. Following the same methodology, we group tasks by textbook series, grades, and learning units (Figure 2). A fine-grained analysis on the distribution of tasks per learning unit within each textbook is presented in Appendix A, Figure 6.

Based on the assumption that a gradual progression would result in multi-task exercises to contain verbs assigned to consecutive Bloom levels, we aggregate these multi-task exercises in four categories: (1) low level + high level, (2) low level + middle level, (3) middle level + high level and (4) any other combination, most likely consisting of all levels. While the number of such tasks is modest (15%), this reveals another concerning trend, once again confirming a general lack of pedagogical principles. Except for *ArsLibri*, all textbooks across all grades mix low levels and high levels in more than half of all multi-task exercises. Fur-

thermore, *ArtKlett, Litera* and *Intuitext* exhibit this trend consistently across most grades and chapters.

If we analyze just one example of a task from a textbook designed for the 5th grade (*ArtKlett*), we can justify several challenges: „Select four verbs from the given text and write them on your worksheet. Use them in a composition in which you present a story set in the garden, showing how you make friends with one of the creatures in the text. In your composition, you will fulfill the following requirements: (1) use the four verbs, with the possibility of changing their form; (2) present a story from the garden, showing how you make friends with a creature from the text; (3) give an appropriate title to your composition; (4) follow the structure of a composition: introduction, contents and conclusion; (5) write your essay in at least ten lines; (6) you will write correctly, using correct phrasing, punctuation and layout.". The assignment asks students to use four verbs in a creative context, which requires higher-order thinking skills such as synthesis (*creating* a story) and *evaluation* (reflecting on their experiences). However, the initial focus on simply selecting and using verbs (*understand, apply*) may lead to confusion about the primary cognitive demand. While Bloom's Taxonomy emphasizes distinct cognitive levels, from remembering to creating, this task blurs those lines by combining various levels without clear differentiation. Moreover, the requirement to "make friends" with a creature introduces an imaginative aspect that can be challenging to evaluate and assess at a cognitive level within Bloom's Taxonomy. The subjective nature of forming friendships makes it difficult to measure students' cognitive engagement and understanding effectively. Last, but not least, while structure is essential for effective writing, the task's emphasis on format (introduction - content - conclusion) may detract from the creative process. A more explicit delineation of the cognitive demands of the instructional tasks, coupled with a more focused pedagogical approach, would facilitate a better achievement of the intended learning outcomes.

5 Discussion and Conclusions

The present analysis provides a systematic view of how Romanian secondary school textbooks promote critical thinking through task design, as operationalized via Bloom's Taxonomy. This responds directly to the concerns raised in prior research (Paul and Elder, 2007; Facione, 1990; Halpern, 1998) and discussed in the literature review, where critical thinking was framed as a key educational competence often underrepresented in instructional materials. While previous work by Chitez et al. (2024) focused on the linguistic and structural complexity of these textbooks, highlighting issues such as lexical overload, syntactic density, and redundancy, our current findings complement that perspective by offering a cognitive-level analysis of instructional tasks. Together, both dimensions reveal a misalignment between task complexity and learner accessibility: even when textbooks attempt to include higher-order tasks (*analyze*, *evaluate*, *create*), these are often embedded in overly dense or poorly scaffolded materials, which may negatively impact rather than support competence development (**RQ1**).

To address **RQ2** and **RQ3**, our results show that only a few textbook series, most notably *Intuitext*, demonstrate a structured progression in cognitive demands across both chapters and grades, reflecting an intentional effort to build students' reasoning skills over time. In contrast, other series, such as *ArtKlett*, while displaying a large variety of tasks, distribute higher-order activities uniformly, lacking clear instructional sequencing. This confirms earlier concerns (Chitez et al., 2024) that task overload and lack of cognitive pacing may dilute the intended pedagogical impact. Moreover, several series (*Corint*, *Litera*, *Paralela45*, *Aramis*) display static cognitive profiles dominated by lower-order tasks, further reinforcing a pattern of surface-level engagement already observed in their linguistic structures.

These findings reinforce arguments from the literature that effective textbook design must not only include a range of cognitive operations but must also organize them in a way that reflects developmental progression (Anderson and Krathwohl, 2001; Graesser et al., 2011). Critical thinking cannot be effectively sustained by simply including complex verbs or isolated higher-order tasks. These should be embedded within a deliberate instructional sequence and supported by clear, accessible language. For textbook designers, this implies a double perspective: to align task design with students' cognitive growth and to calibrate language complexity to ensure engagement and understanding.

The combined model of cognitive and linguistic evaluation offers a replicable framework for

assessing instructional quality in educational materials. Future research could extend this methodology by integrating argumentation mining, dialogic task analysis, or learner performance data, with the goal of aligning curriculum design more closely with evidence-based models of competence development.

Finally, the annotated set of Bloom Taxonomy labels for verbs in instructional materials developed in this study provides a valuable resource for NLP research in education. By linking specific verb forms to cognitive processes such as *remember*, *analyze*, or *create*, this dataset enables more granular and automated assessments of instructional intent. Unlike traditional readability metrics, which focus on surface-level linguistic features, Bloom-aligned annotations allow for the identification of pedagogical depth and cognitive demand. This enables a range of novel applications, including automatic task classification, educational question generation, and curriculum alignment modeling, particularly in low-resource educational settings like Romanian. Moreover, integrating Bloom-labeled data into NLP models can enhance the interpretability of text complexity predictions and support the development of AI-driven tools for textbook evaluation, instructional design, and adaptive learning systems.

Limitations

While this study effectively identifies and categorizes the most frequent verbs associated with Bloom's Taxonomy within the ROTEX corpus, it has limitations that must be acknowledged. From a pedagogical standpoint, focusing solely on verbs does not provide a complete picture of the instructional tasks and learning objectives. The meaning and effectiveness of verbs can vary significantly depending on the context in which they are used. A verb might imply different cognitive processes based on the surrounding tasks, objectives, and instructional strategies. Therefore, the rating and categorization process was highly contextual in the case of ROTEX.

In this context, one limitation of this study is that automated keyword matching alone cannot capture the full nuance of instructional language, often resulting in oversimplified cognitive labels. To address this, we applied a contextual annotation strategy that considers how verbs function within the broader pedagogical framing of each task. This method improves the accuracy of classification by accounting for cases where the intended cognitive process is not clearly expressed through verbs alone. This variation underscores the need for an interpretive approach, and our annotation method responds by capturing how Bloom's Taxonomy operates in real instructional contexts, where taxonomic intent is often implicit.

A separate limitation is the absence of formal inter-annotator agreement scores in the Bloom Taxonomy label annotation process. Although we did not compute quantitative measures of annotation reliability, we addressed annotation reliability by implementing a multi-phase validation pipeline involving trained student raters and subsequent expert review. This layered approach, particularly the involvement of domain experts in the final verification, helped ensure that annotations aligned closely with pedagogical intent and contextual usage. Similarly, while the assignment of a single Bloom Taxonomy level per task, based on the first verb, might appear reductive given the complexity of some prompts, this simplification was necessary for large-scale processing and related automatic disambiguation procedures. The automation is further supported by an additional sentence-level analysis of multi-task exercises. Such choices balance methodological rigor with the practical demands of corpus-level annotation.

Furthermore, although the ROTEX corpus focuses exclusively on Romanian Language and Literature textbooks and includes uneven grade-level representation for some publishers, it still captures the full range of instructional materials currently in use. Finally, while our study does not include learner performance data, it offers a strong foundation for future work linking task design with educational outcomes. By combining linguistic analysis with pedagogical classification, the article effectively contributes a resource and methodology that can inform both textbook development and automated curriculum evaluation.

Acknowledgments

This research is supported by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021-2027, MySMIS no. 334906 and CNCS/CCCDI UEFISCDI, SiRoLa project, PN-IV-P1-PCE-2023-1701, within PNCDI IV, Romania.

References

Laura K Allen, Erica L Snow, and Danielle S McNamara. 2015. Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. In *Proceedings of the fifth international conference on learning analytics and knowledge*, LAK '15, pages 246–254, New York, NY, USA. Association for Computing Machinery.

Lorin W Anderson and David R Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition*. Addison Wesley Longman, Inc.

Gary Beauchamp and Steve Kennewell. 2010. Interactivity in the classroom and its impact on learning. *Computers & Education*, 54(3):759–766. Learning in Digital Worlds: Selected Contributions from the CAL 09 Conference.

Benjamin S Bloom, Max D Engelhart, Edward J Furst, Walker H Hill, David R Krathwohl, and 1 others. 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. Longman New York.

Stephen D Brookfield. 2011. *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. John Wiley & Sons.

Gamze Çavdar and Sue Doe. 2012. Learning through writing: Teaching critical thinking skills in writing assignments. *PS: Political Science & Politics*, 45(2):298–306.

Madalina Chitez and 1 others. 2024. Linguistic overload in secondary school textbooks: A corpus-informed case study of Romanian 6th grade textbooks. In *Conference Proceedings. Innovation in Language Learning 2024*.

Scott A Crossley and Danielle S McNamara. 2016. Text-based recall and extra-textual generations resulting from simplified and authentic texts. *Reading in a Foreign Language*, 28:1–19.

John Dewey. 2022. *How we think*. DigiCat.

Robert H Ennis. 1985. A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2):44–48.

Peter Facione. 1990. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). *Research Findings and Recommendations*, 315.

Arthur C Graesser, Danielle S McNamara, and Jonna M Kulikowich. 2011. Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5):223–234.

Roger Graves. 2017. *Writing assignments across university disciplines*. Trafford Publishing.

Diane F Halpern. 1998. Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4):449.

Diane F Halpern. 2013. *Thought and knowledge: An introduction to critical thinking*. Psychology Press.

Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

Deanna Kuhn. 2005. *Education for thinking*. Harvard University Press.

Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The impact of generative ai on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–22.

Ryo Miyazaki. 2024. Using revised Bloom's taxonomy to evaluate higher order thinking skills (HOTS) in tasks from an 8th grade English language textbook in Japan. Master's thesis, Southeast Missouri State University.

Maryam Mizbani, Hadi Salehi, Omid Tabatabaei, and Mohammadreza Talebinejad. 2023. Textbook evaluation based on Bloom's revised taxonomy: Iranian senior high school textbook in focus. *Language and Translation*, 13(1):85–99.

Gerald M. Nosich. 2005. *Learning to think things through: A guide to critical thinking across the curriculum (2nd ed.)*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Alexandru Oraviţan, Mădălina Chitez, and Roxana Rogobete. 2023. A linguistically-informed assessment model for multidimensional competence building in Romanian school writing. *Educatia 21*, pages 85–91.

Richard Paul and Linda Elder. 2007. Critical thinking: The art of Socratic questioning. *Journal of Developmental Education*, 31(1):36.

David Perkins. 2008. *Smart schools: From training memories to educating minds*. Simon and Schuster.

A Mehdi Riazi and Narjes Mosalanejad. 2010. Evaluation of learning objectives in Iranian high-school and pre-university English textbooks using Bloom's taxonomy. *Tesl-Ej*, 13(4).

John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Harry Torrance. 2007. Assessment as learning? how the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3):281–294.

Jeroen JG Van Merrienboer and John Sweller. 2005. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17:147–177.

Lev Semenovich Vygotsky and Michael Cole. 1978. *Mind in society: Development of higher psychological processes*. Harvard University Press.

A Additional details

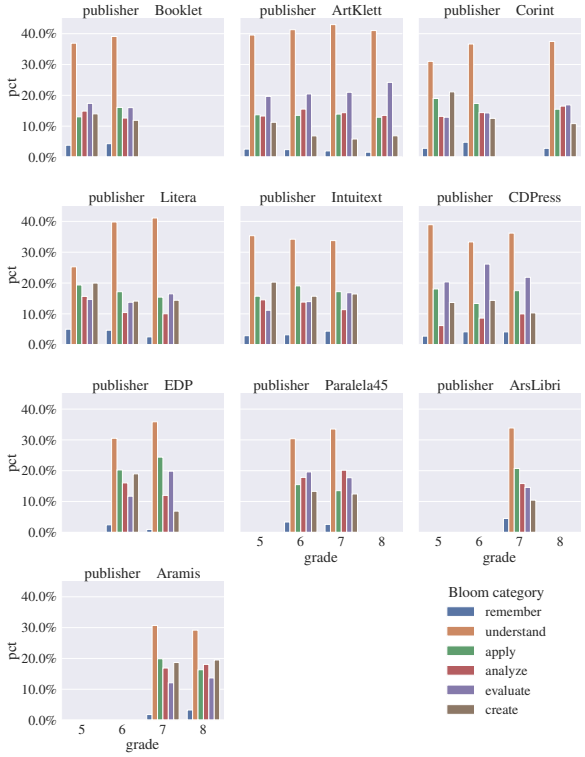

Figure 3: Bloom category percentage counts by publisher and grade

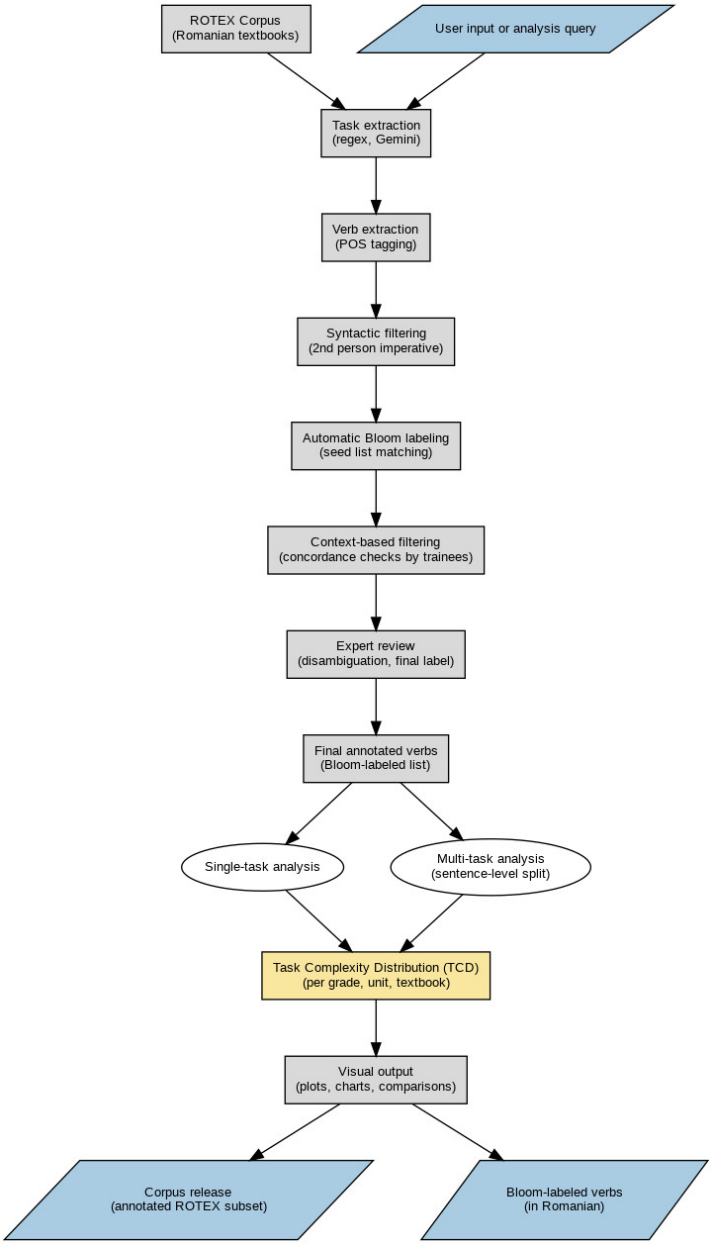

Figure 4: Pipeline for annotating Bloom's Taxonomy labels

Gemini Prompt
You are given a page from a Romanian language and literature textbook written in Romanian. Extract all exercises from the page. Usually, exercises are numbered. Keep the number of exercises in the structured output. Do not shorten or summarize the text of the exercises. Use the full text that is presented in the document. Do not remove any sentence from the exercise text. Keep the order of the exercises as they appear on the page. Unite the syllabified words in the exercises. Make a JSON file of the output. The output JSON should include all the exercises from the file with their full text. Include the page number and the name of the section that each exercise belongs to in the JSON file. The section name is typically found just before the exercises. If the section name is not provided on the page, leave that field empty. Use an integer to indicate the order of the sections in the document; the first section should be labeled as 1, the second as 2, and so on. There can be multiple sections with the same name in the document. If the document does not contain any exercise, leave the JSON file empty.

Table 4: Prompt used for extracting tasks from textbooks

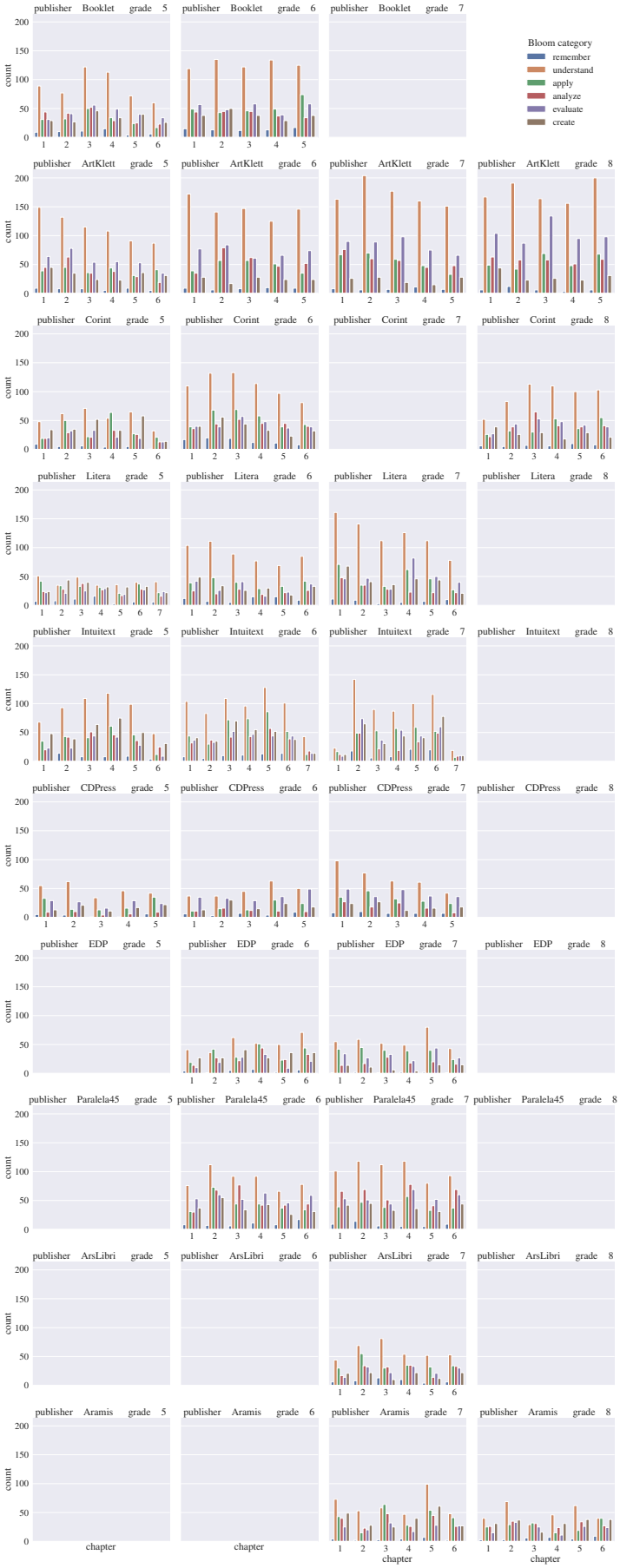

Figure 5: Bloom category counts by publisher, grade, and chapter

Figure 6: Bloom category counts for multitask exercises by publisher, grade, and chapter