

**Multimodal Representation of Academic Anxiety
in Jory John's The Smart Cookie**

**Nabila Zuki¹, Sarah Safika Sitorus², Ibnati Faiza Hamdi³,
Anggun Dhanisa⁴, Muhammad Yusuf⁵**

^{1,2,3,4,5} Universitas Sumatera Utara - Indonesia

Email: nabilazuki721@gmail.com¹, sarahsafikasitorus@gmail.com²,
ibnatifaiza24@gmail.com³, anggundhanisa@gmail.com⁴, yusuf_my@usu.ac.id⁵

Abstract

Article History:

Received: 18/06/2026

Accepted: 06/07/2026

Published: 10/07/2026

Keywords:

*Multimodal discourse
analysis, Picture books,
Academic anxiety,
Social-emotional
literacy, The Smart
Cookie*

Modern children's literature no longer serves merely as a tool for early literacy, but functions as a socio-cultural artifact reflecting the emotional and psychological challenges of contemporary children. However, previous research on how children's academic anxiety and identity crises that are deeply depicted through the synergy of text and imagery in popular storybooks remains highly under-explored. This study aims to examine the multimodal construction in the picture book *The Smart Cookie* by Jory John and Pete Oswald (2021), specifically how the protagonist's "baked identity" serves as a layered semiotic metaphor for perfectionist pressures and academic performance anxiety. Methodologically, this study adopts a qualitative descriptive approach within the Multimodal Discourse Analysis (MDA) paradigm, integrating Michael Halliday's Systemic Functional Linguistics (SFL), Kress and van Leeuwen's visual social semiotics, and intermodal relations. The analysis focuses on the verbal, visual, spatial, and gestural dimensions of 18 selected double-page spreads. The findings demonstrate that the text-image dynamics actively construct a "third story," a psychological realm arising from the ironic tension and counterpoint between verbiage and illustration. This study concludes that micro-level narrative and visual decisions successfully encode macro-level social anxieties, allowing the storybook to critique academic performativity while simultaneously fostering internal psychological resilience and social-emotional literacy in young readers.

Kata Kunci:

*Analisis Wacana
Multimodal, Buku
Bergambar,
Kecemasan Akademik,
Sastra Anak, The Smart
Cookie*

Abstrak

Sastra anak modern saat ini tidak lagi sekadar menjadi alat literasi dini, melainkan berfungsi sebagai artefak sosio-kultural yang merefleksikan tantangan emosional dan psikologis anak kontemporer. Meskipun demikian, penelitian terdahulu mengenai bagaimana kecemasan akademik dan krisis identitas anak digambarkan secara

mendalam melalui sinergi teks dan gambar dalam buku cerita populer masih sangat terbatas. Penelitian ini bertujuan untuk menguji konstruksi multimodal dalam buku cerita bergambar *The Smart Cookie* karya Jory John dan Pete Oswald (2021), khususnya bagaimana identitas “kue” sang protagonis menjadi metafora semiotik berlapis untuk tekanan perfeksionisme dan kecemasan performa akademik. Penelitian ini menerapkan pendekatan deskriptif kualitatif dalam paradigma Analisis Wacana Multimodal (MDA) dengan mengintegrasikan Linguistik Sistemik Fungsional (SFL) Michael Halliday, semiotika sosial visual Kress dan van Leeuwen, serta hubungan intermodal. Analisis difokuskan pada dimensi verbal, visual, spasial, dan gestural dari 18 ilustrasi halaman ganda (*double-page spreads*) terpilih. Temuan penelitian menunjukkan bahwa dinamika wacana teks-gambar secara aktif membangun “cerita ketiga” (*third story*), yaitu ruang psikologis yang muncul dari ketegangan ironis dan kontrapoin antara elemen teks (*verbiage*) dan ilustrasi visual. Penelitian ini menyimpulkan bahwa keputusan naratif dan visual pada tingkat mikro berhasil mengodekan kecemasan sosial tingkat makro, sehingga buku cerita ini mampu mengkritik performativitas akademik sekaligus menumbuhkan resiliensi psikologis internal dan literasi sosial-emosional pada pembaca muda.

INTRODUCTION

Children's literature is no longer viewed merely as a simple tool for early literacy; rather, it is understood as a complex socio-cultural artifact that reflects, questions, and shapes the psychological and emotional realities of human life (Hunt, 2005; Reynolds, 2011). It possesses a unique capacity to encode macro-social anxieties into accessible narrative frameworks, thereby serving as a cultural mirror that helps young readers navigate identity, societal expectations, and personal agency (Nodelman, 2008). Consequently, the content of contemporary children's literature has increasingly focused on highlighting complex internal conflicts, such as psychological distress, the pressure of perfectionism, emotional fragility, and internalized anxiety experienced by individuals living in an increasingly competitive social world (Ko, 2025). Today, children's stories do not merely offer didactic moral lessons; instead, modern narratives within this realm actively

negotiate the intricate emotional landscapes of academic stress and cognitive dissonance, while mapping the complex processes of self-discovery and resilience (Guijarro, 2016).

To convey these profound narratives, modern children's picture books rely on the interplay of two essential, collaborative elements: written text and illustrative imagery. Because meaning in a picture book is co-constructed by both text and illustration, the two cannot be separated during analysis (Nikolajeva & Scott, 2006). Therefore, this study employs a Multimodal Discourse Analysis (MDA) approach, focusing on four distinct dimensions: visual, verbal, spatial, and gesture analysis. This MDA method helps elucidate how text and imagery work in tandem to construct the emotional atmosphere of a story and influence the reader's emotions (Painter, Martin, & Unsworth, 2013; Huang & Xu, 2025).

A considerable body of previous research has utilized this multimodal method to dissect the relationship between text and image in children's media. For instance, prior studies have demonstrated that this theory is highly effective in portraying a character's perspective and explaining how psychological events are represented visually ("Reading Visual Narratives: Image Analysis of Children's Picture Books," 2013; Yefymenko, 2024). Other studies have shown how word choices and illustrative details at a foundational level can reflect broader social conditions (Shi, 2023). Extensive research has focused on text–image interaction from perspectives primarily dominated by compositional, decorative, or visual aspects. However, a significant analytical gap remains. Limited attention has been given to how multimodal resources working in an integrated manner represent a child's academic anxiety, the process of identity formation, and psychological resilience in contemporary picture books, as clearly illustrated in *The Smart Cookie*.

To address this gap, this study focuses on analyzing a picture book titled "*The Smart Cookie*" (2021) by Jory John and Pete Oswald, a New York Times bestselling author and illustrator team. The story centers on a cookie character in a bakery who feels unintelligent at school, struggles to keep up with lessons, and suffers from low self-esteem, until it ultimately discovers its own creativity and intelligence through a poetry writing assignment (John & Oswald, 2021). Within

this narrative, the protagonist's identity as a "cookie" is not merely a cute character design, but rather a symbol of the anxiety and academic pressure frequently experienced by children.

The primary objective of this study is to examine how the textual and visual structures in *The Smart Cookie* work synergistically to depict the protagonist's emotional journey. Through this analysis, we aim to uncover how linguistic and visual meanings are constructed on each page, while mapping the character's transformation from initially feeling "lost at sea" to ultimately realizing that everyone can be "smart in many different ways" (John & Oswald, 2021).

The significance of this study is twofold. Academically, it contributes to the development of Multimodal Discourse Analysis by demonstrating how visual, verbal, spatial, and gesture analyses can be integrated to understand psychological states in children's literature. Practically, this research helps educators, literary critics, and parents better understand how contemporary children's books can validate children's feelings and emotional struggles. The findings of this study are highly expected to foster healthier discussions regarding the diversity of intelligence, academic pressure in children, and the vital importance of safeguarding children's mental health from an early age.

Literature review

Over the past few decades, there has been a significant paradigm shift in the academic discourse surrounding children's picture books, with a growing significance on viewing the medium as a dense, multi-layered site for meaning-making where semiotic choices go beyond simple linguistic representation or decorative illustration (Moya-Guijarro & Ventola, 2023). Recent research has further clarified the crucial role of these dual semiotic systems in coordinating interpersonal relationships and social affiliations between readers, characters, and overarching thematic content within this macro-level theoretical landscape (García, 2021, ; Lirola, 2020). In particular, these studies highlight how the strong synergy between verbal and visual modes serves as an essential pedagogical scaffold, introducing young readers to highly prized domains of cultural ideology, literary

interpretation, and emotional cognition (Wang & Yin, 2024). Evaluative meanings are the exact structural mechanisms through which attitudes, judgments, and internal emotional states are encoded. They are a fundamental, but historically under-examined, part of how stories are made in visual media (Shi, 2020).

As we going deeper into the deductive triangle concerning character construction, researchers have emphasized that visual characters serve as essential vehicles for fostering social-emotional literacy, offering young readers a secure, metaphorical distance to explore their intricate psychological growth and internal struggles (Tisnawijaya & Kurniati, 2024). Despite this acknowledgment, a persistent disparity characterizes the existing research: studies often emphasize the static compositional arrangement of visual elements or the formal structures of art rather than the nuanced, dynamic evaluative functions of the accompanying text (Dieu, 2021; Jumaa & Ali, 2025). Thus, the real-time interaction between verbal and visual semiotic resources is significantly underexplored, especially concerning how language selections in the text may actively limit, ground, or enhance the perception of dynamic visual character attributes (Qiu, 2019).

Rectifying this significant analytical deficiency necessitates a more stringent, methodical implementation of Systemic Functional Linguistics (SFL) and social semiotic theory to ascertain precisely how verbal and visual modalities collaboratively operate in shaping a character's emotional destiny. This study seeks to integrate Michael Halliday's Systemic Functional Grammar with Gunther Kress and Theo van Leeuwen's visual social semiotics, mapping their correspondences as a unified entity rather than as separate communicative channels (Litzler, 2017; Painter et al., 2013). This integrated approach enables a profound analysis of how attitudinal meanings are co-instantiated to encode both convergence (where text and picture mutually reinforce) and divergence (where text and image contest each other) between textual and pictorial systems (Hu & Qiu, 2019). This research recognizes the interplay between visual and verbal semiotic modes in fostering social inclusion and emotional awareness, resulting in intersemiotically coherent multimodal texts that reverberate across several psychological dimensions (Guijarro & Ventola, 2021; Royce, 2002).

This study employs this framework to analyze the intricate "third story" produced by the narrative, a psychological realm arising from the ironic tension or incongruity between text and image, which provides a more profound and subversive critique of contemporary social pressures than either medium could accomplish alone (Yuqiao, 2025). This paradigm elucidates the intersection of evaluative verbal patterns, encompassing the explicit coding of affect, judgment, and appreciation, with visual vector processes to shape the protagonist's internal psychological trajectory (Mills et al., 2020; Rajah, 2022). This study reveals that the protagonist's narrative development is not simply presented to a passive audience, but is actively performed through the deliberate coordination of verbal evaluation and visual character emphasis (Unsworth, 2011).

Notwithstanding these theoretical advancements, a significant, unresolved analytical gap endures at the foundational level of our investigation: the exact word-by-image mechanisms of this intersemiotic synergy are insufficiently theorized, especially regarding the internal, unseen psychological landscapes of fictional characters undergoing contemporary stress. Current multimodal discourse analyses often excessively depend on either discrete visual compositional components (such as framing or color schemes) or overarching linguistic patterns, inadequately examining the nuanced, co-existing evaluative meanings that emerge from the visual emphasis (Dieu, 2021; Jumaa & Ali, 2025). This study rectifies the imbalance by conducting a focused and systematic examination of the text-image dynamics in *The Smart Cookie*, thereby reconciling the historical disjunction between superficial narrative description and the structural analysis of character-driven psychological narratives (Qiu, 2019).

This study examines the simultaneous occurrence of attitudinal lexicogrammar and visual semiotic modes, demonstrating how characters' internal states are fundamentally expressed through the integration of these distinct yet converging semiotic streams (Kilby & Lennon, 2021; Kopatich et al., 2019, p. 12). This thorough amalgamation of verbal and visual elements demonstrates how *The*

Smart Cookie adeptly manages the performative demands of academic success, skillfully converting the playful bakery environment into a diagnostic arena for contemporary educational anxiety (Cohn, 2014, p. 1; Wildfeuer, 2019). This study examines the intricate multimodal patterns to explore the profound ideological ramifications of character focalization, illustrating how micro-narrative techniques mirror, uphold, and occasionally challenge overarching sociocultural standards related to identity, value, and emotional management in a state of anxiety (Forceville, 2023, p. 180; Jabeen, 2022).

METHODS

This research adopts a qualitative descriptive approach within a Multimodal Discourse Analysis (MDA) framework." (Creswell, 2018). This method is used to systematically evaluate and interpret both the verbal and visual modes taken from a children's picture book within the sampled figure and text that make up the data. The verbal data is made up of linguistic units such as words, phrases, and clauses, whereas the visual data is made up of graphic elements like illustrations, layouts, and color schemes. The Smart Cookie (2021), a picture book by Jory John and Pete Oswald, is the main source of data, with particular attention paid to important double-page spreads that chart the protagonist's psychological development. A various theoretical approach combining Visual Social Semiotics (Kress & van Leeuwen, 2021), the Multimodal Narrative model (Painter et al., 2013), and Intermodal Relations (Martinec & Salway, 2005) was used to thoroughly analyze the data.

The data were gathered using a documentation approach combined with purposive sampling. Purposive sampling involves selecting sample units based on specific criteria relevant to the research objectives (Creswell, 2012). In this study, the criterion of the data acquired is words and figures that incorporate social and cultural materials, specifically those charting the protagonist's psychological development. The statistics are obtained from existing material, which is the English picture book. The primary data source, The Smart Cookie (2021), consists of a total 30 pages. For the purpose of this study, 18 specific pages were selected

and comprehensively analyzed based on their explicit representation of the narrative's core psychological shifts. The sample was selected based on two main thematic criteria: (1) representations of the protagonist's academic anxiety and psychological distress, and (2) visual or verbal indicators of emotional resilience and self-actualization. By isolating these 18 critical pages rather than evaluating the entire book, this study ensures a highly focused operational mapping of the character's psychological trajectory from conflict to resolution without irrelevant narrative data.

After identifying and categorizing the data, the data went through a comprehensive multimodal analysis. This study uses a blended theoretical approach, combining Visual Social Semiotics (Kress & van Leeuwen, 2021), the Multimodal Narrative model (Painter et al., 2013), and Intermodal Relations (Martinec & Salway, 2005). The analysis is executed through four stages based on the multimodality framework of Kress and van Leeuwen: verbal analysis, visual analysis, spatial analysis. The linguistic structures (words, phrases, clauses, sentences), textual functions and meanings are used to identify the data in the form of text during the verbal analysis step. The visual elements (such as objects, characters, backdrops, colours) and visual meanings (which analyse the symbolisms and meanings that may be read through visual elements) are used to identify the data in the form of visual images during the visual analysis stage. The layout and spatial structure are used to identify how visual elements are arranged in relation to one another across the pages during the spatial analysis stage. The portrayed characters' body language, facial expressions, and posture are used to identify and interpret psychological and narrative shifts during the gesture analysis stage.

Data validation in this qualitative study was achieved through theoretical responsiveness and systematic intra-rater consistency. Rather than relying on subjective interpretation, the analysis strictly adhered to the explicit structural criteria and semiotic rules established by Kress and van Leeuwen (2021), Painter et al. (2013), and Martinec and Salway (2005). Reliability was further ensured through iterative rounds of analysis, wherein the researcher consistently cross-checked early

analytical findings against later data to guarantee that the verbal, visual, spatial, and gestural frameworks were applied uniformly across the entire dataset.

FINDINGS AND DISCUSSION

The analysis demonstrates how contemporary children's literature critiques academic performativity and constructs young characters' internal psychological resilience. The characters themselves are based entirely on Western-style bakeries and use a metaphorical world described as a bakery environment and school housed in a gingerbread house. The picture book, entitled “The Smart Cookie”, written by Jory John and illustrated by Pete Oswald, encompasses a dynamic narrative arc across critical double-page spreads. Each selected spread in the book deploys textual and visual elements that delineate abstract emotional states, self-worth, and shifting conceptualizations of intelligence. This article divides the analysis into three main categories: verbal mode, visual mode, and multimodal interaction analysis.

4.1 Verbal Analysis

Dialogue, lexical choice, and narrative voice shape how intelligence and failure are framed for young readers. Figure 1 shows the past experience of the main character, which was filled with self-doubt and alienation. The dark-colored background adds to the sad atmosphere that explains the character's position as separated from the other group as the other food is inside and the main character is alone at the outside of the bakery, plus the sad expression that the character makes as they stare at their reflection in the puddle visually represents the feeling of not being accepted. Meanwhile, the text explains that the character finds it difficult to express ideas and does not consider themselves smart. This relationship between the visual and verbal reinforces the theme of self-doubt, which becomes the main conflict in this story.

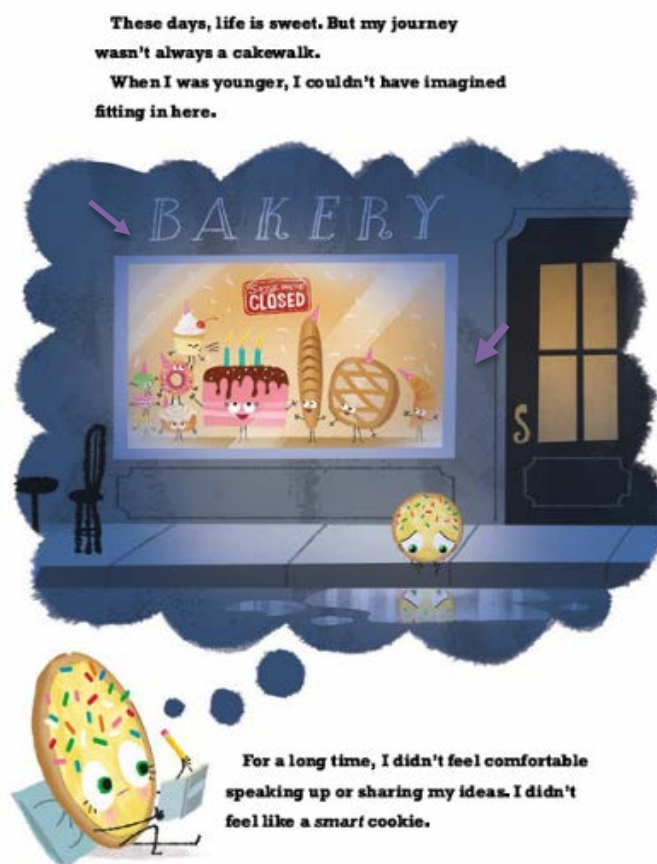


Figure 1. Multimodal Analysis of *[The smart cookie]* – page

The first-person flashback narration in Figure 2's vocal part establishes the emotional foundation of the story Data 1: "These days, life is sweet. But my journey wasn't always a cakewalk... I didn't feel like a smart cookie." From a linguistic standpoint, the author uses the pronoun "I" in a confessional tone to establish intimacy and empathy with the young readers right away. Moreover, the deliberate incorporation of idioms related to food, such as "cakewalk" to indicate a simple task and "smart cookie" to indicate intelligence, serves two purposes. It deftly maintains the book's light-hearted bakery premise while also clearly expressing the weighty, abstract mental health issues of social estrangement and academic self-doubt.

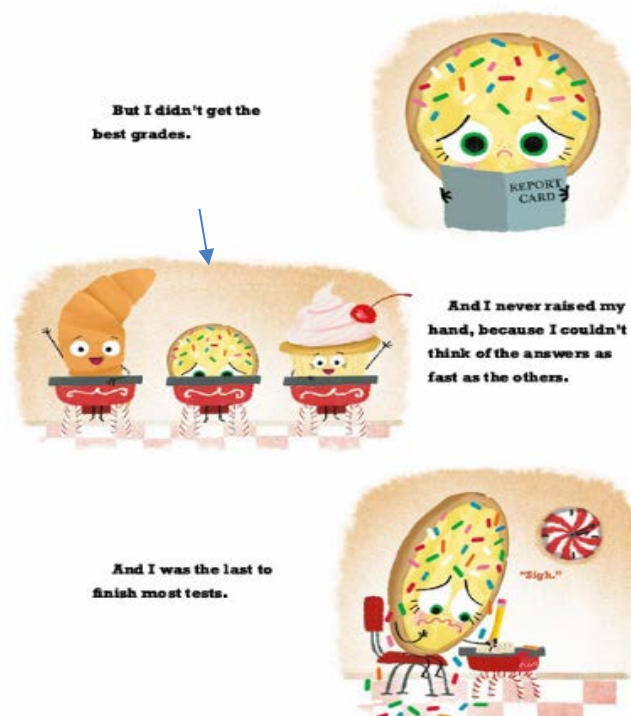


Figure 2. Multimodal Analysis of [The smart cookie] – page

Data 2: The text blocks describing the fast-paced nature of the classroom tasks: *"The teacher was always speaking fast. My homework was always a mess. I was always the last one to finish."* This section makes use of structural parallelism and the adverb "always" in three separate phrases. This grammatical pattern modifies the story's tempo to reflect the character's constant, unrelenting pressure at school. The character's internal fear is strongly connected to the perceived overwhelming speed of the academic environment, as seen by the repeated sentence structure that vocally supports a chronic state of helplessness rather than occasional failure. In Figure 2, the character states the struggle of not being the best, as other cakes always thinks faster and finish the test earlier than the main character that results the main character's grades not being the best. This proves the main character's struggle with pressure.

Data 3: The textual repetition of the time-phrase "*day after day after DAY*", with the final word rendered in all-capital letters. Typographical manipulation is used on this page to increase the story's emotional stakes. The language changes from a passive description of time to an active linguistic representation of mental tiredness by capitalizing the last word. In addition to concluding with a subtle textual cliffhanger that forces the reader to anticipate a significant turning point in the cookie's psychological trip, this language exaggeration successfully conveys the character's ongoing uneasiness.



Figure 3. Multimodal Analysis of [The smart cookie] – page

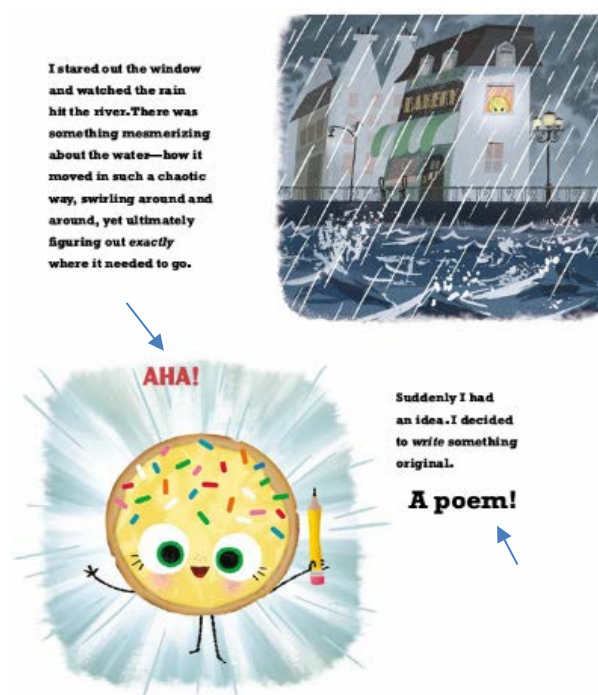


Figure 4. Multimodal Analysis of *[The smart cookie]* – page

Data 4: The sudden introduction of bolded, capitalized exclamations and short sentences: "AHA!" and "A poem!" These particular lexical elements function as indicators of abrupt cognitive enlightenment in discourse analysis, dramatically shifting the book's previously depressing tone. Linguistically, the character's reclaiming of agency is indicated by the transition from lengthy, passive statements of concern to brief, animated exclamations. The reader is prepared for the character's transformation from a nervous outsider to a self-assured creative by this abrupt shift in textual energy.

4.2 Visual Analysis

Visual analysis, which looks at non-linguistic semiotic resources that create mood and thematic depth, takes over in the second stage. This section deciphers how color schemes, lighting, backgrounds, and symbolic pictures generate characters' psychological moods. It is based on Tafrijiyah (2025) multimodal method. By materializing abstract ideas like academic anxiety, failure, and security,

these visual components serve as semiotic tools that enable young readers to understand emotional weight without only depending on text. Color, perspective, and composition are interpreted as deliberate meaning-making tools influencing children's perceptions of intelligence and resilience, drawing on Kress and van Leeuwen's Visual Grammar (2006). The illustrator of *The Smart Cookie* externalizes Cookie's transition from self-doubt to creative confidence through the use of warm versus cool tones, lighting contrasts, and compositional focus.

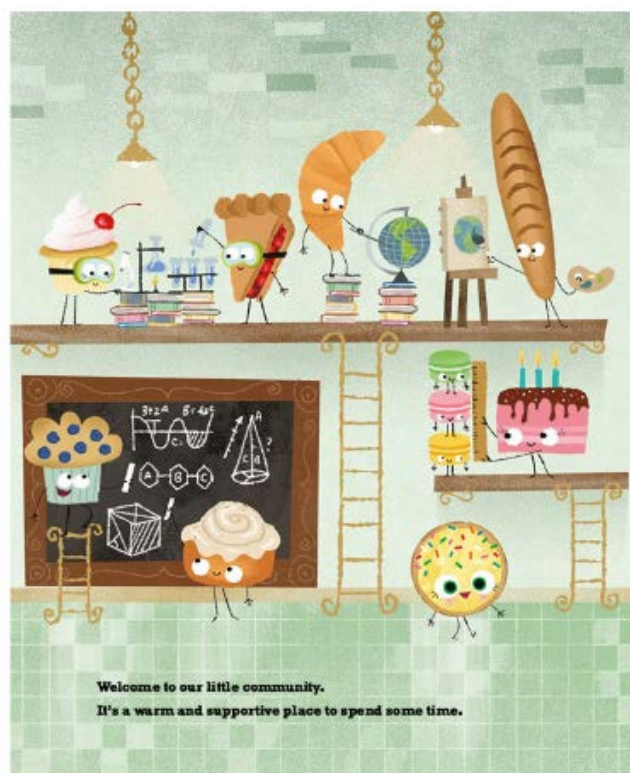


Figure 5. Multimodal Analysis of *[The smart cookie]* – page

Brightly colored cookie characters are displayed in the visual composition against a gentle pastel-green tiled background that evokes a tidy and orderly safe educational setting. Warm pastel colors serve as a visual aid in multimodal discourse analysis that communicates coziness, security, and approachability, mirroring socially accepted ideas of schools as encouraging learning environments. Structure, stability, and discipline are also indicated by the clean background and

well-organized graphic components. Additionally, the contrast between the colorful characters and the subdued background both highlights the characters and places them in an instructive context. Before the story focuses on the protagonist's emotional struggles, these visual elements work together to provide a good portrayal of the school setting, creating a sense of safety and possible development.

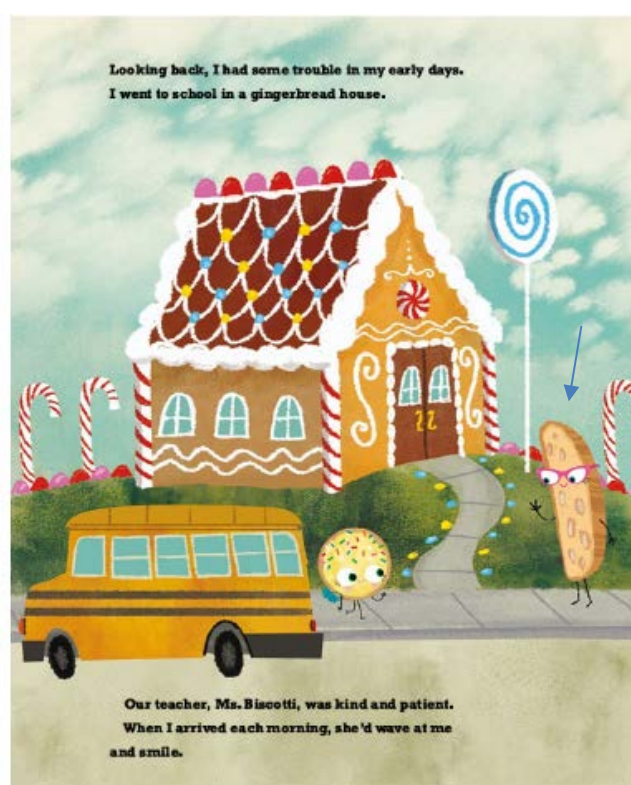
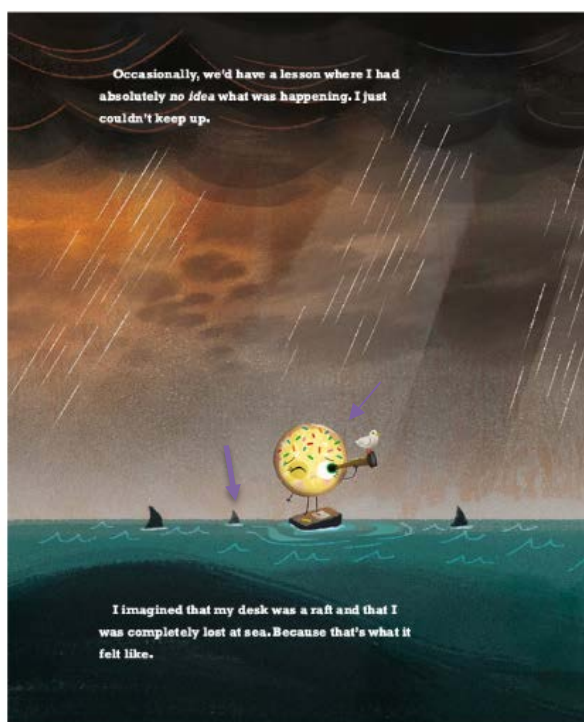


Figure 6. Multimodal Analysis of [The smart cookie] – page

An illustration of a whimsical gingerbread house school under a bright, clear sky, alongside the welcoming and smiling figure of the teacher, Ms. Biscotti. The image employs cheerful and vibrant colors to construct the school as a welcoming and supportive environment rather than a source of fear or anxiety. Within the visual framework, the gentle facial expressions, soft contours, and approachable character design of Ms. Biscotti function as semiotic resources that communicate warmth, care, and trustworthiness. Her visual prominence in the composition positions her as a symbolic representation of adult guidance and institutional support, reinforcing the role of educators as figures of protection and reassurance. Through these visual

elements, the image promotes a positive perception of the school environment and highlights the presence of supportive relationships within the educational setting.



Picture 7. Multimodal Analysis of [The smart cookie] – page

Dark, ominous rainclouds, a turbulent sea, and sharp shark fins circling a tiny school desk where the cookie stands with a telescope. The main character is surrounded by sharks in the middle of the ocean. Visually describe the main character as a vulnerable target in an overwhelming and unsafe environment. This page demonstrates a strong use of visual symbolism by transforming an abstract emotional experience into a concrete and easily recognizable visual representation. The dominant dark grey and deep blue color palette creates a somber atmosphere associated with fear, anxiety, and emotional distress. The imagery of shark fins and a turbulent storm functions as visual metaphors for the overwhelming nature of academic self-doubt, uncertainty, and psychological pressure. These elements position the cookie character within a threatening environment, visually representing feelings of helplessness and confusion. Through the interaction of color, symbolism, and composition, the image effectively externalizes the

protagonist's internal emotional struggle, allowing viewers to perceive the intensity of the character's anxiety in a tangible form.

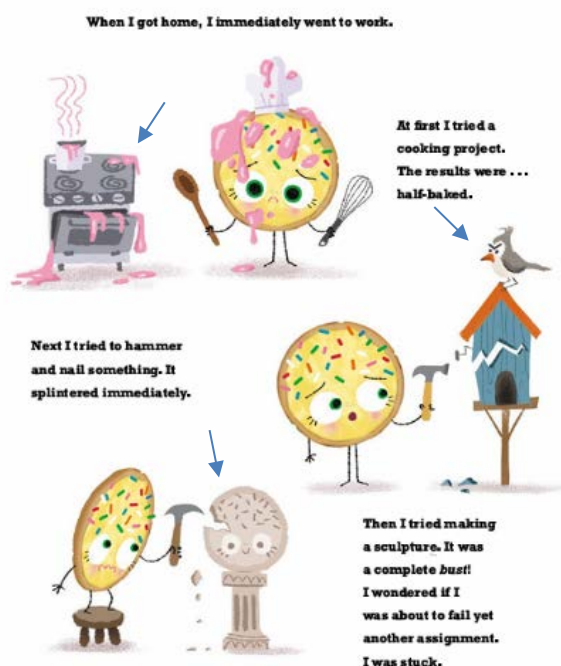


Figure 8. Multimodal Analysis of [The smart cookie] – page

Broken artifacts are lined up against a stark white backdrop in a minimalist display of a smoldering oven, a collapsed birdhouse, and a cracked sculpture. This describes that the character tried to do many things to find out what the main character is good at but he failed as the cooking is ruined, the birdhouse is splintered, and the sculpture is broken. By contrasting these destroyed endeavors with a blank canvas, the illustrator removes any intimidating mood, transforming failure into something playful and funny. The literal cracks and smoke serve as direct visual signifiers of "mistakes," making the concept of trial and error relatable and un-intimidating for children.

4.3 Spatial Analysis

The third phase of the analysis involves Spatial Analysis, which examines the layout, compositional arrangement, and organizational boundaries of the pages. Following Tafrijiyah (2025) framework, this section dissects how the positioning

of text and images (such as center versus margin) and the manipulation of physical space affect the reader's interpretation. Spatial analysis illustrates how the deliberate confinement or liberation of visual elements within the page layout can physically elicit sensations of restriction, isolation, or ultimate freedom, mirroring the main character's psychological condition through the page's geometry.

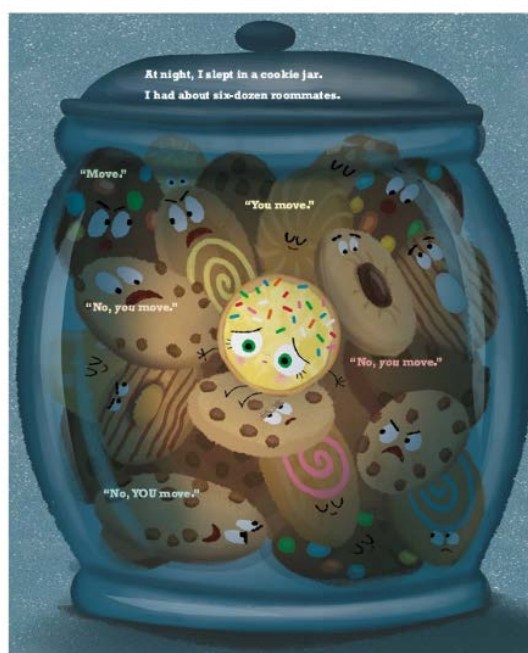


Figure 9. Multimodal Analysis of [The smart cookie] – page

The multimodal discourse analysis of figure 9, the spatial layout and positioning serve as the primary semiotic resources to convey intense psychological confinement. Compositionally, the main sprinkle cookie is placed directly in the Center of the frame, surrounded and encircled by tiny dialogue text blocks like "Move" and "You move." Rather than showing dominance, this central positioning highlights the character's vulnerability, as the surrounding text acts as a circular, inescapable wall of social friction.

Furthermore, the central cookie is tightly packed and squeezed by dozens of other cookies. This dense crowding visually minimizes the protagonist's scale, conveying that they feel incredibly small, helpless, and completely stuck (stuck) in a deadlock with no way out or room for change. Trapped inside the rigid circular lines of the massive blue glass cookie jar, the entire composition eliminates negative

space. This lack of breathing room physically restricts the characters, transforming the literal jar into a powerful visual metaphor for claustrophobia, social discomfort, and total existential stagnation.

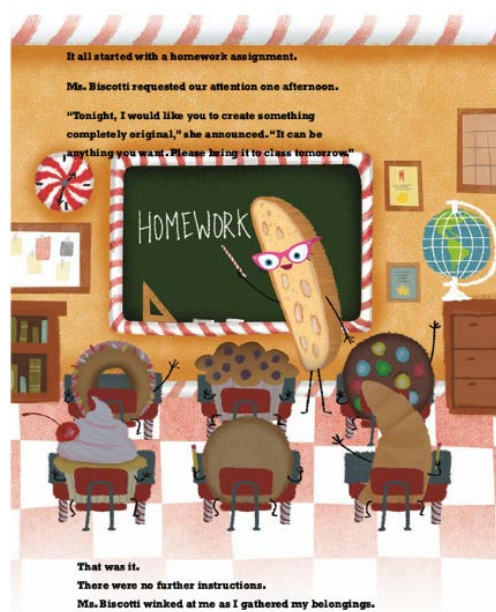


Figure 10. Multimodal Analysis of [The smart cookie] – page

Figure 10's composition uses a strict centripetal structure to support institutional hierarchy and power dynamics in the classroom, according to the multimodal spatial analysis. Strong leading lines are created by the symmetrical arrangement of the pupils' desks, drawing the viewer's attention inward to the precise Center, in which Ms. Biscotti and the chalkboard are located. This dead-center positioning establishes the instructor as the principal knowledge mediator in visual grammar by giving her complete salience and institutional authority. The student characters' back-facing perspective, which produces an alignment of gaze (vector of attention), emphasizes this centralized focus even more. The reader is drawn into the same controlled, subservient attitude as the classroom participants, gazing upward toward the center of power, since the audience is viewing the scene from behind the kids.

Moreover, a distinct ideological division of labor and institutional conformance are mapped by the spatial distribution between the foreground and

background. The lower part of the frame is made up of neat desk rows that symbolize the organized, disciplined setting of traditional education. The assignment is framed as an unavoidable focus point in the upper portion, which is dominated by the large blackboard with the phrase "HOMEWORK" printed on it. Ms. Biscotti's stature is physically elevated above the children due to her towering stature and vertical extension toward the upper edge of the chalkboard. The physical and geometric space of the classroom is purposefully designed to demand obedience, order, and complete attention toward institutional demands. This spatial distance and height differential visually represent the power gap between the authoritative educator and the uniformed students.



Figure 11. Multimodal Analysis of [The smart cookie] – page

The horizontal structural division in figure 11 employs Kress and van Leeuwen's Ideal/Real compositional framework to illustrate the protagonist's emotional transition in the multimodal spatial analysis. The upper zone (the "Ideal") illustrates a dynamic, exuberant playground environment focused on a slide, symbolizing a communal, idealized condition of joyous youthful interaction and external expression. In contrast, the lower zone (the "Real") anchors the narrative in a more sober, pragmatic reality: a personal, interpersonal exchange in which Ms. Biscotti presents the protagonist with a private message of encouragement.

This horizontal separation is further augmented by the deliberate application of vertical framing and negative space. The broad white gutter dividing the two scenes serves as a structural "visual pause," facilitating a cognitive shift between the chaotic, public exterior of the playground and the introspective realm of the lower frame. Moreover, the arrangement of the typographical parts connects this spatial void. The text blocks are intentionally organized in a descending staircase configuration along the right margin, creating a distinct visual trajectory that guides the reader's gaze downward. This musical trajectory reflects the protagonist's transition in focus from the chaotic distractions of exterior playground activities and converging into a profoundly personal, contemplative state prompted by a serene moment of affirmation.



Figure 12. Multimodal Analysis of [The smart cookie] – page

The multimodal spatial analysis of figure 12 reveals a transition from conventional institutional hierarchy to a centrifugal, non-linear arrangement that honors pluralism and human autonomy. The layout features five diverse drawings of the cookie figure arranged in a dynamic circular orbit around an open white core, deliberately eschewing the conventional grid format of rows and columns commonly linked to standardization. In visual grammar, this decentralized configuration dismantles the singular authority characteristic of earlier classroom arrangements, redirecting emphasis towards decentralized, self-directed forms of identification.

The significant use of negative space serves as an essential semiotic resource across this site. The surrounding white background functions as a vast buffer zone, providing each vignette with autonomy and prominence rather than indicating isolation or emptiness. This allows each individual expression of talent, whether it is filmmaking, skateboarding, weightlifting, music, or compassion, to "breathe" independently rather than competing for visual dominance.

Moreover, the amalgamation of word and imagery accentuates the concept of unbounded exploration. The typographical blocks are meticulously isolated at the extreme top and bottom margins, serving as conceptual framing anchors while maintaining an uncluttered center. This configuration guarantees that the open area continues to be a dynamic realm of possibilities. The visual vector of the layout does not impose a strict chronological order on the reader's gaze; rather, it promotes a multi-directional reading experience. This particular geographical configuration reflects the fundamental educational principle of the linguistic text: that intellect and self-actualization are complex, unbounded, and intrinsically varied.

4.4 Gesture Analysis

The concluding part of the multimodal framework is Gesture Analysis, which investigates the conveyance of meaning through non-verbal communication, encompassing facial expressions, gaze, body position, and various forms of kinesic behavior. In **The Smart Cookie**, movements are integral to the emotional and

psychological evolution of the protagonist, enhancing the verbal narrative and visual composition. In contrast to traditional instructional materials that typically emphasize language content over embodied communication, the book utilizes expressive facial characteristics, eye-level gazes, and dynamic physical postures to convey intricate emotions and interpersonal interactions. The narrative use multimodal cues to graphically depict the Smart Cookie's evolution from self-doubt and vulnerability to confidence and self-acceptance, illustrating how gestures enhance meaning in conjunction with spoken and visual components.



Figure 13. Multimodal Analysis of [The smart cookie] – page

In the multimodal discourse analysis of Figure 13, the focus changes to gesture and kinesic analysis to break down how the main character shows severe stage fright and anxiety. A lot of the main character's visual representation depends on nonverbal somatic cues to show that they are having a mental breakdown. The legs of the cookie are clearly shown with a wavy line that looks like it's shaking. In the language of gesture analysis, this particular lower-limb distortion is a direct

physical expression of physiological panic and somatic anxiety. It makes the feeling of having your knees buckle under extreme fear visible.

This shaky stance is deeply connected to the character's upper-body movements and general posture. The main character's arms are pulled in and tightly gripping a piece of paper close to their chest. This acts as a symbolic barrier between them and the audience, protecting themselves. The character physically shrinks their presence in the area by pulling in and not extending their arms or legs outward. When combined with the slanted eyebrows and wide, sad eyes, these physical choices turn vague feelings of vulnerability and self-doubt into a very clear somatic performance that perfectly supports the written text that talks about shaking hands and being nervous in public.

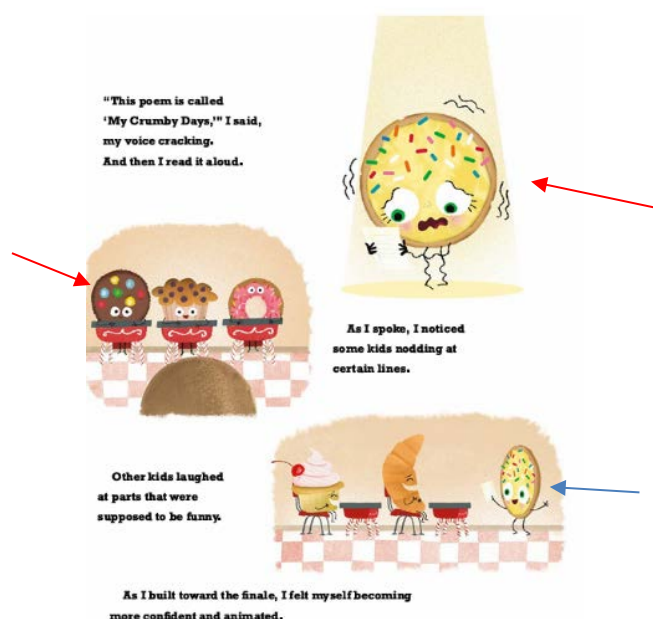


Figure 14. Multimodal Analysis of [The smart cookie] – page

In the gesture analysis of figure 14, the three panels show a clear step-by-step physical transformation from extreme panic to pure confidence. In the first panel, located at the top right, the cookie is frozen in a state of high anxiety. Kinesically, its legs are drawn with shaky, squiggly lines, and vibration marks surround its body to show it is actively trembling, while its face remains tight and defensive with stressed eyes and a wavy mouth that captures total stage fright.

The narrative shifts in the second panel in the middle left, where the focus moves to the audience. Here, the other characters are smiling and nodding, and their welcoming gestures and supportive facial expressions completely change the mood, acting as the trigger that relieves the main character's stress.

Finally, the third panel at the bottom right displays a complete physical release for the cookie. The shaky lines completely disappear, replaced by stable, straight legs and arms thrown wide open in a confident, animated gesture. With wide eyes and a big, warm smile, the character's open posture beautifully shows the exact moment its fear turns into total relief and connection.

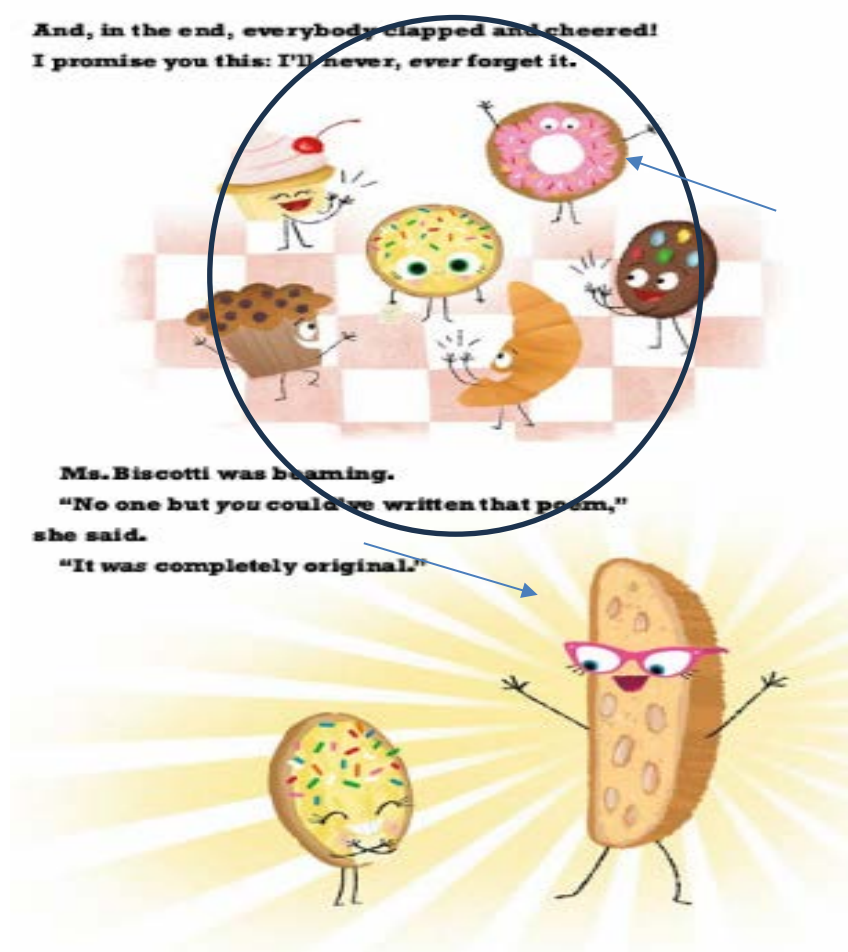


Figure 15. Multimodal Analysis of [The smart cookie] – page

Figure 15's gesture and kinesic analysis shows that all the somatic signs of worry are replaced with big, happy body movements, marking the emotional peak of the visual story. In the bottom picture, Ms. Biscotti and the cookie are shown with their eyes closed and smiles big. There is a bright sunburst design around them that makes their happiness stand out even more. Ms. Biscotti's arms are raised high in the air, making an open, highly expressive motion that means she is completely happy, relieved, and celebrating. This open stance shows institutional pride and full acceptance, which successfully fixes and heals the main character's previous academic weakness.

This kind action directly affects the cookie's posture, which changes to a relaxed, inward-folding hand position close to its chest. It also smiles widely. This kinesic response shows a deep sense of relief and internalized pride. It is very different from earlier scenes, where the person clutched the paper defensively. By coordinating these warm facial emotions and triumphant upper-body extensions, the teacher and student's interaction visually creates a sense of full acceptance and emotional breakthrough, ending the story in a way that is satisfying and reassuring.

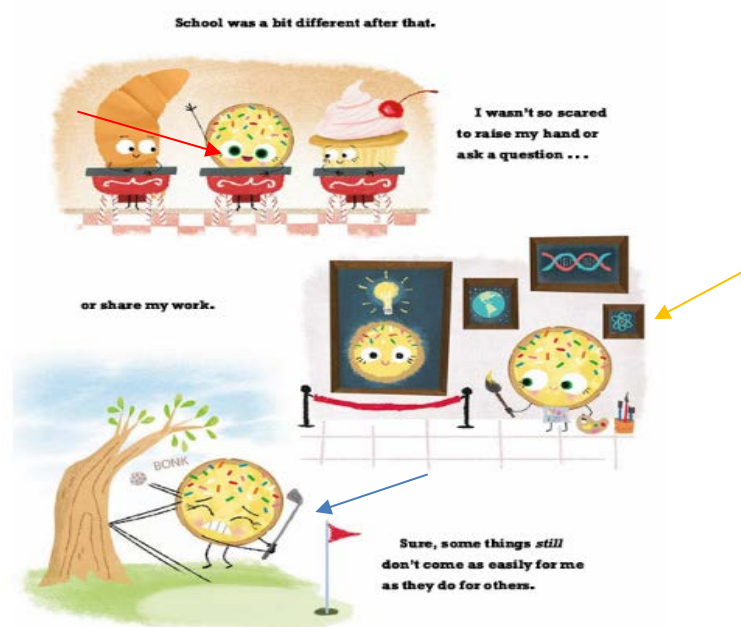


Figure 16. Multimodal Analysis of [The smart cookie] – page

The cookie stands tall with its head held high and its arm raised straight and firmly into the air during a classroom discussion. A powerful sociocultural symbol of academic freedom and involvement is raising one's hand with assurance. The character is no longer cowering in fear but is actively deciding to claim its presence and voice within the learning group, as indicated by the straight posture and direct, upward motion.



Figure 17. Multimodal Analysis of [The smart cookie] – page

A close-up group portrait in which every one of the many bakery characters leans in close, shoulder to shoulder, making eye contact with the reader while grinning. Strong social cohesiveness and a sense of belonging are indicated by the close physical proximity. More significantly, the direct eye-level glance penetrates the narrative wall, drawing the young readers into their circle and delivering an important message of shared identity and acceptance.



Figure 18. Multimodal Analysis of [The smart cookie] – page

The cookie sits triumphantly on top of a tall stack of books, maintaining a relaxed, upright posture and a direct, confident smile aimed straight at the viewer. The ability to overcome one's academic anxieties is symbolized by sitting atop the books. The once-anxious cookie is transformed into an inspiring figure of perseverance who speaks directly to the reader's own potential through the use of a steady, direct stare and an open, relaxed stance as a last act of empowerment.

Discussions

Together, the results show that The Smart Cookie creates meaning through the methodical interaction of four semiotic layers: verbal, visual, spatial, and gestural. These layers work together rather than separately to tell a story. This supports Kress and van Leeuwen's (2006) and Painter et al.'s (2013) claim that picturebook meaning is co-instantiated across modes; however, the present analysis extends this position by showing precisely how convergence and divergence between modes produce what Yuqiao (2025) terms a "third story." Where prior studies (e.g., Shi, 2020; Tafrijiyah, 2025) have tended to isolate either the verbal-evaluative or the visual-compositional dimension, this study's four-stage model demonstrates that the psychological interiority of the protagonist becomes legible

only when verbal appraisal, visual symbolism, spatial confinement, and embodied gesture are read as a single coordinated system.

The change from the early to later parts of the story best illustrates this cooperation. Hu and Qiu's (2019) claim that attitudinal meaning is amplified when modes reinforce rather than merely accompany one another is consistent with the convergence of a passive, hedge-laden verbal register (Data 2, Data 3), a visually darkened and threat-saturated palette (Figure 7), a centripetal, boundary-eliminating spatial layout (Figures 9–10), and somatic markers of panic (Figure 13). Conversely, the later spreads (Figures 12, 14–18) demonstrate mode divergence functioning as resolution: verbal restraint ("You just need a chance to try...") paired with visually decentralized, high-agency compositions produces a calmer, more distributed form of affect than either mode achieves alone. This provides empirical evidence for Yuqiao's (2025) theoretical assertion that ironic or counterpoint relations between text and image produce interpretive surplus; however, this study also demonstrates that the same principle applies to convergent (non-ironic) relations, indicating that the concept of the "third story" may need to be expanded beyond incongruity alone.

The results of the study also address the gap in the literature review of the "internal, unseen psychological landscapes" of characters (Qiu, 2019; Dieu, 2021; Jumaa & Ali, 2025). The rigid circularity of the cookie jar (Figure 9) and the centripetal hierarchy of the classroom (Figure 10) are not neutral backdrops but rather active constructors of the character's felt confinement. This study shows that spatial geometry itself serves as a direct index of psychological state, whereas previous multimodal research has primarily addressed spatial and compositional analysis as concerned with formal arrangement. This refines the theoretical framework put out by Hu and Qiu (2019) and implies that spatial semiotics should be seen as an evaluative resource on par with verbal assessment rather than as a secondary, compositional-only issue.

At a broader level, these findings position *The Smart Cookie* as doing ideological work that exceeds its surface function as an early-literacy text. By resolving the protagonist's arc not through improved academic performance but through a redefinition of intelligence itself (Figure 12's decentralized, non-hierarchical layout), the book multimodally enacts a critique of standardized achievement narratives, resonating with Takahashi's (2013) discussion of multiple-intelligence frameworks in inclusive education. This positions the text's pedagogical value not merely in modeling resilience, but in modeling a structural alternative to performativity — a distinction with implications for how educators select and use picturebooks addressing academic anxiety, and one that moves this study's contribution beyond description into argument: multimodal resources in children's literature can function as sites of ideological negotiation, not just emotional illustration.

CONCLUSION

This study demonstrates that *The Smart Cookie* functions as a socially and culturally constructed text that effectively mirrors the internal psychological struggles of contemporary children. By integrating systemic functional linguistics, visual grammar, spatial layout, and gesture analysis, the narrative constructs a powerful “third story” where textual and visual modes function together to critique rigid academic performativity. The manuscript illustrates a clear semiotic shift from systemic confinement and self-doubt to decentralized, creative agency, ultimately offering young readers a comprehensive view of intelligence. The supportive role of the educator underscores the vital impact of academic validation on a child's internal resilience. These micro-level narrative choices serve a macro-level educational purpose, transforming a simple children picture storybook into an essential tool for fostering early childhood social-emotional literacy and psychological well being.

Despite its comprehensive textual insights, this study has several limitations. First, the analysis is limited to a single case study, which is a one specific children's picture book (*The Smart Cookie*). So, the findings cannot be

generalized to all contemporary children's literature that talking or dealing with academic anxiety. Second, because this research is a study adopts a qualitative descriptive approach within the Multimodal Discourse Analysis (MDA), it lacks of factual data regarding how children, parents, or teachers actually read, interpret, and emotionally respond to the book's multimodal ideas. Based on the limitations, future researchers could expand the scope by conducting a comparative multimodal analysis across Jory John's entire Food Group series (such as *The Bad Seed* and *The Good Egg*) or other modern picture books addressing mental health themes to identify broader visual and verbal trends. Future studies could also embrace field research, such as interviews or classroom observations, to look at how children actually decode these multimodal elements and whether the book successfully reduces their academic performance anxiety.

REFERENCES

- Cohn, N. (2014). The architecture of visual narrative comprehension: The interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5, Article 680. <https://doi.org/10.3389/fpsyg.2014.00680>
- Dieu, H. N. T. (2021). An introduction into using the grammar of visual design framework an evaluative theory in analysing the interface between pictures and words in picture books for children. *Journal of Science Social Science*, 66(3), 101–108. <https://doi.org/10.18173/2354-1067.2021-0052>
- Forceville, C. (2023). Narrating and focalizing visually and visual-verbally in comics and graphic novels. *Pragmatics & Cognition*, 30(1), 180–208. <https://doi.org/10.1075/pc.22007.for>
- García, C. S. (2021). A Semiotic and Multimodal Analysis of Interactive Relations in Picture Books That Challenge Female Gender Stereotypes 1. In *Routledge eBooks* (pp. 144–163). Informa. <https://doi.org/10.4324/9781003145875-10>
- Guijarro, A. J. M., & Ventola, E. (2021). Picture Books, Gender and Multimodality. In *Routledge eBooks* (pp. 1–20). Informa. <https://doi.org/10.4324/9781003145875-1>
- Guijarro, J. M. (2016). *Multimodal Analysis of Children's Literature*.
- Hu, Y., & Qiu, Q. (2019). A Social Semiotic Approach to the Attitudinal Meanings in Multimodal Texts. *Theory and Practice in Language Studies*, 9(9), 1160–1160. <https://doi.org/10.17507/tpls.0909.12>
- Huang, D., & Xu, J. (2025). Extending local grammar to multimodal meaning-making: A case study of conflict resolution in picture books. *Text & Talk*, 46(2), 167–197. <https://doi.org/10.1515/text-2024-0095>
- Hunt, P. (2005). *Understanding Children's Literature*. Routledge.

- Jabeen, M. (2022). Multimodal storytelling: Exploring the semiotic patterns of narration and focalization in “Pakistan Girl.” *Pakistan Languages and Humanities Review*, 6. [https://doi.org/10.47205/plhr.2022\(6-ii\)108](https://doi.org/10.47205/plhr.2022(6-ii)108)
- John, J., & Oswald, P. (2021). *The Smart Cookie*. HarperCollins.
- Jumaa, M. Z., & Ali, H. S. (2025). A Multimodal Discourse Analysis of the Compositional Meaning of Moral Lessons in Picture Books for Children. *JOURNAL OF LANGUAGE STUDIES*, 9, 357–378. <https://doi.org/10.25130/lang.9.4.p2.18>
- Kilby, L., & Lennon, H. (2021). When words are not enough: Combined textual and visual multimodal analysis as a Critical Discursive Psychology undertaking. *Methods in Psychology*, 5, 100071. <https://doi.org/10.1016/j.metip.2021.100071>
- Ko, C.-C., Chiang, C.-H., Lin, Y. L., & Chen, M.-C. (2025). Multimodal children's e-books help young learners in reading.
- Kopatich, R. D., Feller, D., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research Principles and Implications*, 4(1). <https://doi.org/10.1186/s41235-019-0176-1>
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge. <https://doi.org/10.4324/9781003099857>
- Lirola, M. M. (2020). A multimodal analysis of Daddy's Roommate : deconstructing compositional and interpersonal meanings. *Brno Studies in English*, 2, 25–46. <https://doi.org/10.5817/bse2020-2-2>
- Litzler, M. F. (2017). A multimodal analysis of picture books for children: A systemic functional approach. *Language Value*, 8(1). <https://doi.org/10.6035/languagev.2016.8.5>
- Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication*, 4(3), 337–371. <https://doi.org/10.1177/1470357205055928>
- Mills, K. A., Stone, B. G., Unsworth, L., & Friend, L. (2020). Multimodal language of attitude in digital composition. *Written Communication*, 37(2), 135–166. <https://doi.org/10.1177/0741088319897978>
- Moya-Guijarro, J., & Ventola, E. (2023). Analysing Picture Books that Challenge Gender Stereotypes Multimodally. *Revija Za Elementarno Izobraževanje*, 16, 9–26. <https://doi.org/10.18690/rei.16.spec.iss.2983>
- Nikolajeva, M., & Scott, C. (2006). *How Picturebooks Work*. Routledge.
- Nodelman, P. (2008). *The Hidden Adult: Defining Children's Literature*. Johns Hopkins University Press.
- Painter, C., Martin, J. R., & Unsworth, L. (2013). *Reading the Visual in Children's Picture Books*. Equinox eBooks Publishing, 1–13. <https://doi.org/10.1558/equinox.18627>
- Qiu, Q. (2019). A Multimodal Analysis of the Interplay between Visual and Verbal Semiotics in Creating Messages in Chinese Picture Books. *Studies in Literature and Language*, 19(3), 44–50. <https://doi.org/10.3968/11444>

- Rajah, S. K. S. (2022). A multimodal analysis of selected literary elements in award-winning Malaysian picturebooks / S. Komathy Senathy Rajah. In University of Malaya Students Repository.
- Reading visual narratives: image analysis of children's picture books. (2013). *Choice Reviews Online*, 50(12), 50–6602. <https://doi.org/10.5860/choice.50-6602>
- Reynolds, K. (2011). *Children's Literature: A Very Short Introduction*. Oxford University Press.
- Royce, T. (2002). Multimodality in the TESOL Classroom: Exploring Visual-Verbal Synergy. *TESOL Quarterly*, 36(2), 191–191. <https://doi.org/10.2307/3588330>
- Shi, Y. (2020). Appraising in visual narratives: a multimodal discourse analysis of Jimmy Liao's picturebook stories. PolyU Institutional Research Archive (Hong Kong Polytechnic University). <http://hdl.handle.net/10397/88241>
- Shi, Y. (2023). Appraising in visual narratives: A multimodal discourse analysis of picturebook stories. *International Journal of Multimodal Discourse Studies*, 11(3), 45–58.
- Tafrijiyah, K. (2025). Multimodal analysis of verbal and visual elements in *The Day You Begin*. *Linguistics and Literature Journal*, 6(2), 172–180. https://doi.org/10.33365/linguistics_and_literature.v6i2.990
- Takahashi, J. (2013). Multiple intelligence theory can help promote inclusive education for children with intellectual disabilities and developmental disorders: Historical reviews of intelligence theory, measurement methods, and suggestions for inclusive education. *Creative Education*, 4(9), 605–610. <https://doi.org/10.4236/ce.2013.49086>
- Tisnawijaya, C., & Kurniati, G. (2024). Anna Kang's Picture Books: Inculcating Young Minds with Social-Emotional Literacy. *Prosodi*, 18(2), 306–317. <https://doi.org/10.21107/prosodi.v18i2.26594>
- Unsworth, L. (2011). Exploring a “good enough” description of image-language interaction for multimodal comprehension and composition pedagogy. In *RUNE (Research UNE)*. <https://hdl.handle.net/1959.11/10670>
- Wang, J., & Yin, J. (2024). The Interpersonal Meaning Analysis of Children's Language-Learning-Focused Picture Books: From a Multimodal Discourse Analysis Perspective. *OALib*, 11(6), 1–19. <https://doi.org/10.4236/oalib.1111659>
- Wildfeuer, J. (2019). The Inferential Semantics of Comics. *Poetics Today*, 40(2), 215–234. <https://doi.org/10.1215/03335372-7298522>
- Yefymenko, V. (2024). Multimodal text-image synergy in representing interpersonal relations in picture books. *Cognition Communication Discourse*, 28, 102–109. <https://doi.org/10.26565/2218-2926-2024-28-07>
- Yuqiao, Z. (2025). The Third Story: An Exploration of Integrating Ironic Counterpoint into the Multimodal Text-image Relationship of Children's Picturebooks. In *University of the Arts London Research Online (University of the Arts London)*. University of the Arts London. <https://ualresearchonline.arts.ac.uk/id/eprint/24875/1/Final%20Thesis%20Submission%20Form%2018%20sep.docx>