

DIDAKTIKA

Jurnal Pendidikan Sekolah Dasar

Volume 9, Nomor 1, 101–114, 2026

Journal homepage: <https://journal.uny.ac.id/index.php/didaktika>



The Cognitive–Affective Asymmetry of Animated Video in Elementary Character Education

Dody Filado^{1✉}, Haryanto², Troy³

¹²Instructional Technology, Yogyakarta State University

Affiliation Address: Jl. Colombo No. 1, Karangmalang, Caturtunggal, Depok, Sleman, 55281, Indonesia, ³Silchar, Assam, India

³Animation Study Program, Institut Seni Indonesia Yogyakarta

Affiliation Address: Jl. Parangtritis KM 6.5, Glondong, Panggungharjo, Sewon, Bantul, 55188, Indonesia

✉Corresponding Author: dodyfilado.2017@student.uny.ac.id

Abstract

Indonesian development research on animated video for character education routinely reports effectiveness for both cognitive and affective outcomes. Most such conclusions, however, rest on one-group pretest-posttest designs and self-report instruments vulnerable to testing, ceiling, and social desirability effects. The present study used the prevailing template deliberately, to examine the gap between what it can detect and what it claims. An animated video on rights, obligations, and responsibility, built with the Alessi and Trollip model, was tested with seventeen fifth-grade students at a public elementary school in Kepanjen, Malang. The cognitive test produced a normalised gain of 0.37, classified as medium but constrained by an already high pretest baseline. The responsibility self-report shifted by only two percentage points, well within measurement noise, with both measurements compressed against the very high category. This asymmetry between a moderate cognitive gain and a negligible affective shift is not idiosyncratic; it is what short multimedia interventions tend to produce when affective outcomes are read as dispositional traits rather than process variables. Animated video supported knowing but, within this study, produced no comparable movement in becoming. The product itself passed expert validation in the feasible to very feasible range. The asymmetry is better read as a property of the prevailing evaluation template than of the product alone, and concerns what a one-group, self-report design can detect, independent of sample size. For researchers and curriculum developers, affective effectiveness should not be inferred from short single-session designs; designs, instruments, and durations must match the dispositional nature of character.

Keywords: Animated Video, Character Education, Cognitive Theory of Multimedia Learning, Optimism Bias, Pre-Experimental Design

Asimetri Kognitif-Afektif Video Animasi dalam Pendidikan Karakter di Sekolah Dasar

Abstrak

Penelitian pengembangan video animasi untuk pendidikan karakter di Indonesia umumnya melaporkan bahwa media tersebut efektif untuk capaian kognitif maupun afektif. Akan tetapi, sebagian besar simpulan itu bertumpu pada desain satu kelompok pretest-posttest dan instrumen laporan diri yang rentan terhadap efek pengujian, efek langit-langit, dan keinginan sosial. Penelitian ini menggunakan templat dominan tersebut secara sengaja untuk menelaah jarak antara apa yang dapat dideteksi templat itu dan apa yang sesungguhnya diklaim oleh literatur. Mengikuti model pengembangan Alessi dan Trollip, video animasi tentang hak, kewajiban, dan tanggung jawab dikembangkan untuk pembelajaran Pendidikan Kewarganegaraan Kelas V dan diujikan kepada tujuh belas siswa di sebuah sekolah dasar negeri di Kepanjen, Malang. Produk lulus validasi ahli dengan kategori layak hingga sangat layak. Tes kognitif menghasilkan gain ternormalisasi sebesar 0,37 dengan kategori sedang, tetapi terbatas oleh skor pretest yang sudah tinggi. Skor laporan diri tanggung jawab hanya bergeser dua poin persentase, masih dalam rentang derau pengukuran, dengan kedua pengukuran berimpit pada kategori sangat tinggi. Penelitian ini berargumen bahwa asimetri antara perolehan kognitif yang sedang dan pergeseran afektif yang nyaris nol bukanlah kekhasan studi ini,

melainkan justru hasil yang sepatutnya diperkirakan dari intervensi multimedia singkat ketika capaian afektif dibaca sebagai sifat watak alih-alih sebagai variabel proses. Video animasi mampu menopang dimensi mengetahui, tetapi dalam kondisi penelitian ini tidak menunjukkan pergeseran yang sepadan pada dimensi pembentukan watak. Bidang ini membutuhkan desain dan instrumen yang sepadan dengan perbedaan tersebut. Bagi peneliti dan pengembang kurikulum, efektivitas afektif tidak sepatutnya disimpulkan dari desain satu sesi berbasis laporan diri.

Keywords: Bias Optimisme, Desain Pra-Eksperimental, Pendidikan Karakter, Teori Kognitif Pembelajaran Multimedia, Video Animasi

✉ *Correspondent Writer: Dody Filado*
Affiliate Address: Yogyakarta, Indonesia
Email: dodyfilado.2017@student.uny.ac.id

INTRODUCTION

Animated video for character education has become one of the most frequently developed media in Indonesian elementary research. Studies in this area follow a recognisable template. A development team applies the Alessi and Trollip, ADDIE, or 4-D model to produce a video, validates the product through experts, runs a small-group trial, and concludes with a one-group pretest-posttest field test. The conclusion, with rare exceptions, is that the medium is feasible, practical, and effective for both cognitive and affective outcomes. Read individually, each of these studies appears to add a new product to the field. Read collectively, they raise a question that recent international scholarship has increasingly problematised. Either character is more responsive to brief multimedia than the broader literature on moral development would expect ([Brown et al., 2023](#); [Walker & Thoma, 2017](#)), or the dominant evaluation template is detecting something other than character change. The present study takes the second possibility seriously, and uses the same template to expose its limits from inside.

The substantive concern is real. Indonesian Law Number 20 of 2003 on the National Education System defines schooling as the development of both intelligence and noble character, and the Penguatan Pendidikan Karakter programme has placed character at the centre of basic education reform. [Munjiatun \(2018\)](#) documents that classroom practice nevertheless remains heavily verbal, with limited use of structured experiences that engage students emotionally. Internationally, [Arthur et al. \(2016\)](#) frames character education as a long-horizon process integrating intellectual, moral, civic, and performance virtues, while [Brown et al. \(2023\)](#), in a meta-analysis of 214 studies covering more than 300,000 participants, report a small overall effect size of Hedges' $g = 0.24$ for character education, with corrected estimates lower still after adjustment for selection bias. At SDN 2 Curungrejo, where this study was conducted, the homeroom teacher of grade five reported that students could repeat the textbook definition of rights and obligations during oral review yet continued to neglect classroom duties and homework. The gap is not primarily a knowledge gap. It is the gap that [Lickona \(1991\)](#) located between moral knowing on one side and moral feeling and moral action on the other, a tripartite structure that [Purwati et al. \(2024\)](#) recently re-validated as a coherent measurement model in Indonesian secondary classrooms.

This critique has particular force in Indonesian elementary education rather than in educational multimedia research in general. The Penguatan Pendidikan Karakter programme has made character a formal priority of basic schooling, yet, as [Munjiatun \(2018\)](#) documents, classroom practice remains predominantly verbal. At the same time, the development-study template criticised here is the modal genre of Indonesian educational technology research, reproduced across large numbers of undergraduate theses and postgraduate projects that apply the Alessi and Trollip, ADDIE, or 4-D model and conclude with affective effectiveness claims drawn from a single class. The combination of a strong policy mandate, a verbally dominated pedagogy, and a high-volume development-study tradition makes the gap between what the template claims and what it can actually detect a specifically Indonesian methodological problem, not merely a generic concern about instructional video.

The theoretical case for animated video begins with the Cognitive Theory of Multimedia Learning (CTML), now in its most recent restatement by [Mayer \(2024\)](#). Mayer argued that learners process verbal and pictorial information through dual channels of limited capacity, and that meaningful learning emerges when the two channels are coordinated through selection, organisation, and integration. The theory is well evidenced for cognitive outcomes. [Noetel et al. \(2021\)](#), in a systematic review of randomised trials in higher education, reported that adding video to existing teaching produced strong learning benefits with a pooled effect size of $g = 0.80$, while replacing existing teaching with video produced smaller but reliable gains of $g = 0.28$. [Mayer \(2024\)](#) and [Mayer & Fiorella \(2021\)](#) have refined the design principles that produce these effects, including the signalling, segmenting, modality, and pre-training principles. The theory is also explicit about its scope: the dependent variables of interest are knowledge structures, not personality traits. [Moreno \(2006\)](#) extended the framework into the Cognitive-Affective Theory of Learning with Media (CATLM), in which affective and motivational states mediate the cognitive resources a learner invests in a task. The extension is important, but its scope is again process-bound. Affect in CATLM is a state variable such as situational interest, satisfaction, or confusion, not a stabilised disposition. The literature that claims affective effectiveness for short animated videos in producing character change is therefore not so much applying these theories as overreaching them.

Three structural biases compromise the prevailing template, each of which has a substantial empirical literature. The first is the design itself. [Spurlock \(2018\)](#) and [Bierer et al. \(2025\)](#) have argued that the single-group pretest-posttest design cannot, even in principle, support causal claims of intervention effectiveness, because history, maturation, testing, and instrumentation effects cannot be separated from any apparent

treatment effect without a control condition. The second bias is publication bias and effect size inflation. [Ropovik et al. \(2021\)](#) showed that meta-analyses in education routinely fail to adjust for publication bias, leading to inflated effect estimates of unknown magnitude. [Kraft \(2023\)](#) demonstrated that 36% of effect sizes from randomised trials of education interventions on standardised outcomes are below 0.05 standard deviations, and that the published literature systematically overstates the typical effect. [Evans & Yuan \(2022\)](#) found similar patterns in international education studies, with smaller studies showing larger and more variable effects than larger studies. The third bias is social desirability in self-report among children. [Camerini & Schulz \(2018\)](#) demonstrated that children's social desirability bias systematically inflates self-reported social well-being scores, and [Ng et al. \(2025\)](#) recently reported partial measurement invariance of the Children's Social Desirability Scale across grade levels, with younger students reporting higher levels of social desirability than older ones. Self-reported character measures in fifth-grade samples are therefore especially vulnerable to ceiling and desirability effects.

Indonesian research on animated video for character formation can be sorted into three clusters along these methodological dimensions. The first cluster, exemplified by [Widiyasanti & Ayriza \(2018\)](#) and [Wuryanti & Kartowagiran \(2016\)](#), uses multi-school quasi-experimental designs with separate experimental and control groups, longer treatments, and inferential statistics. Studies in this cluster make defensible claims about affective change because the design controls for the most plausible threats to internal validity. The second cluster, into which the majority of recent development studies fall, uses single-class pre-experimental designs with self-report instruments and reports affective effectiveness on the basis of pretest-posttest mean differences alone. Studies in this cluster cannot, by the methodological work cited above, support strong claims about character change, regardless of the magnitude of the reported difference. The third position, where the present study sits, uses the same single-class pre-experimental design as the second cluster but reports cognitive and affective outcomes side by side, and reads any divergence between them as the central finding rather than as a problem to be smoothed over.

The present study makes three contributions. First, it offers a feasible product for an under-served Civic Education topic. Second, it provides preliminary evidence of a moderate but ceiling-bounded cognitive gain alongside a negligible affective shift, in a configuration that allows the two outcomes to be directly compared within the same students. Third, and most importantly, it uses this internal comparison to argue that the dominant evaluation template in the field may systematically bias researchers toward optimistic conclusions about affective effectiveness. The asymmetry reported here is not a methodological accident peculiar to a single school. It is the kind of result that appears likely to recur when short multimedia interventions are evaluated against dispositional outcomes with self-report instruments. This is, in short, an R&D study whose primary scholarly contribution is methodological rather than product-oriented. Throughout, three nested levels are distinguished: character as the broad dispositional construct, responsibility as the specific dimension targeted here, and the self-report responsibility score as its operational measure; claims are calibrated to the level actually measured. In this sense, animated video is *more capable of supporting knowing, insufficient for becoming*, and a research community that wishes to make defensible claims about character formation will need designs, instruments, and intervention durations that match the dispositional nature of the construct it claims to move.

The novelty of this study lies less in the product than in what it does with the dominant evaluation design. Research on animated video for character and civic learning, including recent systematic reviews, overwhelmingly reports positive cognitive and affective effects ([Putri et al., 2024](#)), and the broader body of character-education empirical research, though expanding rapidly, has devoted comparatively little attention to the validity of character assessment itself ([Oldham & McLoughlin, 2026](#)). This study differs by turning the prevailing one-group, self-report template on itself: instead of adding another effectiveness claim, it uses the template deliberately to expose the cognitive-affective asymmetry the template systematically produces, separating what the design can detect (a bounded cognitive gain) from what it cannot (dispositional change). The contribution is therefore methodological—a diagnostic reading of a widely used evaluation pipeline—rather than a new product or a new claim of effectiveness.

METHODS

This study used the research and development approach with the [Alessi & Trollip \(2001\)](#) instructional design model ([Figure 1](#)). The model organises multimedia development into three sequential stages of planning, design, and development, supported by three continuous attributes of standards, ongoing evaluation, and project management. The model was chosen deliberately, in full awareness of its dominance in the literature this study critiques. Adopting the same template makes the results more

directly comparable to prior work and ensures that any divergence between the cognitive and the affective outcome cannot be attributed to a different intervention pipeline.

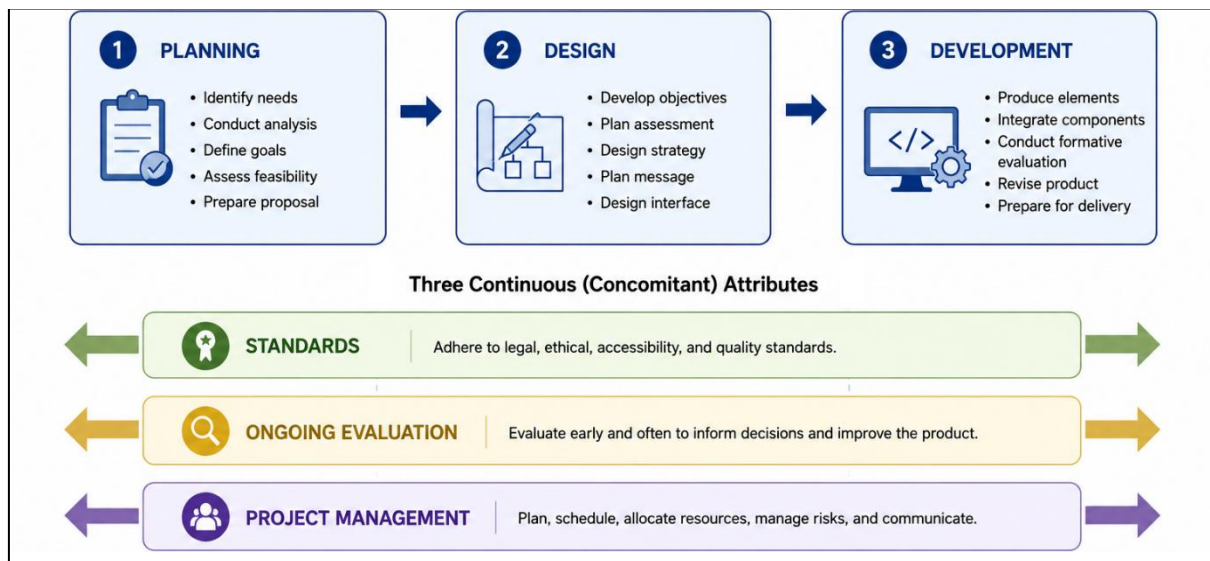


Figure 1. The Alessi and Trollip (2001) Instructional Design Model

In the planning stage, a needs analysis combined classroom observation and a structured interview with the homeroom teacher of grade five at SDN 2 Curungrejo, Kepanjen, Malang. The interview probed the teacher's view of character education, the recurring difficulties faced when teaching rights and obligations, and the media currently available for the topic. A learner characteristics analysis informed decisions on narration pace, vocabulary, character design, and scene length, following the design recommendations of [Mayer \(2021\)](#) for instructional video, particularly the signalling, segmenting, and pre-training principles.

In the design stage, the storyline was divided into four sub-themes that span the daily life of an elementary student, namely rights, obligations, and responsibility as a religious community member, as a student at school, as a child in the family, and as a member of society. The four-domain structure was chosen so that the medium would offer concrete behavioural anchors rather than abstract definitions. Storyboards were produced in three increasing levels of detail, from textual storyboard to visual sketches and finally a digital storyboard ready for animation. Recent reviews of animated video in education by [Kleftodimos \(2024\)](#) and [Mayer & Fiorella \(2021\)](#) informed the production decisions on character design, narration style, and scene length.

In the development stage, digital assets were produced in Adobe Illustrator 2021, narration was recorded in Adobe Audition 2020, and the final assembly was completed in Adobe After Effects 2020. The product was then submitted for alpha testing to one media expert and two content experts, revised, and tested with five students individually before the field trial.

The validation procedure used separate expert instruments for media and for content. The media instrument, completed by one media expert, covered interface design, user control, navigation, information presentation, and help features; the content instrument, completed by two subject-matter experts, covered curricular alignment, material accuracy, and learning design. Each item was rated on a five-point Likert scale and converted to a feasibility category using ideal-mean and ideal-standard-deviation thresholds, and the homeroom teacher completed a parallel practicality rubric. Expert ratings fell in the feasible to very feasible range, with material-expert means of approximately 3.9 to 4.6 out of 5. Only after this validation, together with the small-group trial, was the product advanced to the field trial.

Participants and Field Trial Design

The trial subjects were fifth-grade students at SDN 2 Curungrejo, Kepanjen, Malang. Five students participated in the small-group trial. Seventeen students participated in the field trial, which used a one-group pretest-posttest design conducted across two sessions. The first session administered the cognitive pretest and the responsibility questionnaire as a baseline. The second session, held one week later,

presented the animated video and concluded with the cognitive posttest and the second administration of the responsibility questionnaire.

The seventeen field-trial participants were fifth-grade students aged approximately ten to eleven years, comprising eleven girls and six boys. The study site, SD Negeri 2 Curungrejo (NPSN 20518762), is a state primary school in the village of Curungrejo within the Kepanjen subdistrict, which is the administrative seat of Malang Regency, East Java; the setting is therefore peri-urban rather than remote rural. The school holds the highest national accreditation grade ("A") and is served by mains electricity and fibre-optic internet, indicating a reasonably resourced public school rather than an under-equipped one. These features of the setting bound the transferability of the findings and are returned to in the limitations.

This design has three known threats to internal validity, each of which is named and addressed here rather than treated as a footnote. Repeated exposure to the same cognitive items at pretest and posttest may inflate the posttest score through the testing effect, as documented at length by [Spurlock \(2018\)](#) and [Bierer et al. \(2025\)](#). Maturation across one week may produce small gains attributable to ordinary classroom learning rather than to the intervention. Social desirability is the dominant threat to the affective measure, and [Camerini & Schulz \(2018\)](#) and [Ng et al. \(2025\)](#) have shown that this threat is especially strong among younger primary school children responding to teacher-administered self-reports on socially valued constructs. The present study does not control for these threats. Instead, the threats are used to interpret the results, and to argue that the same threats almost certainly affect the broader literature in which similar designs are dominant.

Instruments and Data Analysis

Four instruments were used. A media expert validation questionnaire covered language, display, usefulness, and closing. A content expert validation questionnaire covered learning aspects and material aspects. A student response questionnaire covered ten items related to clarity, attractiveness, audio and video quality, and perceived helpfulness. A 25-item four-option multiple-choice cognitive test was used at pretest and posttest, alongside a 15-item Likert-scale responsibility questionnaire administered before and after the intervention. The responsibility questionnaire drew on the moral knowing, moral feeling, and moral action domains validated by [Purwati et al. \(2024\)](#) in an Indonesian context. All instruments were reviewed by an instrument expert prior to use.

Validation and response data were analysed by computing item means and converting them to a five-level qualitative scale. Cognitive effectiveness was analysed using the normalised gain proposed by [Hake \(1998\)](#), defined as the ratio of the actual average gain to the maximum possible gain, with thresholds of low for g below 0.30, medium for g between 0.30 and 0.70, and high for g above 0.70. Following [Nissen et al. \(2018\)](#), cognitive change is reported in two complementary forms, namely raw mean change and normalised gain, so that readers can judge for themselves whether the apparent gain reflects genuine learning or a low pretest baseline. Responsibility data were analysed by computing the percentage of total achieved score against the maximum possible score and interpreting the result on a four-level criterion.

RESULTS AND DISCUSSION

Results

The planning stage produced a needs profile in which the teacher reported that no character-oriented media were available for the topic of rights, obligations, and responsibility, and that pandemic-era instruction had relied mainly on synchronous video conferencing and homework distribution. The design and development stages produced an animated video lasting approximately ten minutes, organised in four scenes mapped to the four life domains identified in the design stage.

Expert validation produced the scores summarised in [Table 1](#). The media expert assigned a mean of 4.00, classified as feasible. The first content expert assigned a mean of 4.57, classified as very feasible, while the second content expert assigned a mean of 3.90, classified as feasible. The dominant suggestions across reviewers were to add an explicit opening title, state learning objectives at the start, reposition definitions to the beginning, and include a closing summary. These changes mirror the signalling and segmenting principles described by [Mayer \(2024\)](#) and were incorporated into the revised version.

Table 1. Summary of Expert Validation Results

Validator	Aspects Evaluated	Mean Score	Category
Media expert	Language, display, usefulness, closing	4.00	Feasible
Content expert I	Learning aspects, material aspects	4.57	Very feasible
Content expert II	Learning aspects, material aspects	3.90	Feasible

The small-group trial with five students produced a mean response score of 4.76 across ten items, classified as very feasible. The homeroom teacher's practicality assessment produced a mean of 3.94, classified as feasible. Together with the expert validation, these results establish that the product itself is a defensible artefact: not exceptional, not flawed, recognisably within the quality range of comparable Indonesian-developed character animations. The product thus functions as a vehicle for the study; the principal scholarly contribution lies in the methodological interpretation rather than in the artefact itself. This baseline matters for the analysis that follows. Any asymmetry observed in the field trial cannot be attributed to a poorly designed product.



Figure 2. Documentation of The Field Trial Session at SDN 2 Curungrejo

The field trial ([Figure 2](#)) produced the cognitive results shown in [Table 2](#). The mean pretest score was 74.59, the mean posttest score was 83.88, and the normalised gain was 0.37, which falls within the medium category according to [Hake \(1998\)](#). Two features of the result deserve attention before any effectiveness claim is built on it. First, the pretest mean of 74.59 indicates that students already possessed a substantial portion of the content knowledge before the intervention. The already high baseline mathematically compresses the maximum possible gain, and the medium-category label, while technically accurate, masks the fact that the absolute room for improvement was limited. Second, the same cognitive items were used at pretest and posttest, which means that some unknown portion of the gain reflects familiarity with the items rather than understanding of the content. Read with these two features in view, the cognitive result is best characterised as a moderate but ceiling-bounded gain, not as evidence of strong cognitive effectiveness.

Table 2. Pretest and Posttest Scores in the Field Trial

Variable	Pretest	Posttest
Lowest score	44	50
Highest score	92	100
Mean score	74.59	83.88
Mean difference = 9.29; normalised gain (g) = 0.37, medium category, ceiling-bounded		

The results of the field trial indicate that the developed learning media effectively improved students' learning outcomes. The mean pretest score increased from 74.59 to 83.88 on the posttest, representing a mean gain of 9.29 points. In addition, the lowest score improved from 44 to 50, while the highest score increased from 92 to a perfect score of 100, demonstrating improvement across different levels of student achievement. The normalized gain (g) of 0.37 falls within the medium category (ceiling-bounded), suggesting that the intervention produced a moderate yet meaningful improvement in students' understanding and performance. These findings indicate that the developed media has satisfactory effectiveness in enhancing student learning outcomes during the field trial.

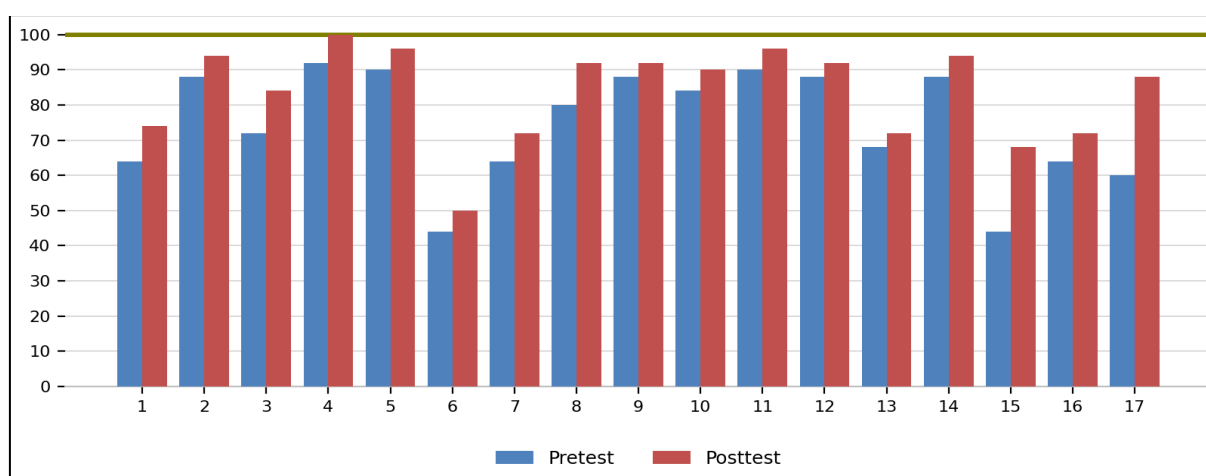


Figure 3. Pretest and Posttest Scores for Each of The 17 Students in The Field Trial

The responsibility self-report produced the results shown in Table 3. The percentage of total achieved score moved from 81% before the intervention to 83% after the intervention, an absolute change of two percentage points across seventeen students. Both measurements fell within the very high category on the four-level criterion, and no student crossed any category boundary (Table 3). This is the signature pattern of a ceiling-compressed self-report described in the measurement literature (Camerini & Schulz, 2018; Ng et al., 2025). Students enter the measurement endorsing items that name socially desirable behaviours, leave very little numerical room for further improvement, and exit the measurement with the same affirmative responses. The change is mathematically positive and substantively negligible. Given the high baseline and the absence of category transition, the observed shift is more plausibly interpreted as measurement fluctuation than as stable behavioural change.

Table 3. Responsibility Score Before and After the Intervention

Time of Measurement	Number of Students	Total Score	Maximum Score	Percentage	Category
Before the intervention	17	1029	1275	81%	Very high
After the intervention	17	1063	1275	83%	Very high

Internal-consistency reliability of the responsibility questionnaire was acceptable (Cronbach's $\alpha = 0.72$ for the fifteen items at the pretest administration). On the 15–75 responsibility-score scale, the scores showed a standard deviation of 5.93 at pretest and 4.32 at posttest, while the cognitive scores showed standard deviations of 15.98 and 13.54 respectively. The compression of the responsibility spread, together with the absence of any category transition, reinforces the reading of the two-point shift as measurement fluctuation rather than dispositional change. Reliability of the cognitive test was not

estimated because its purpose was criterion-referenced achievement assessment rather than psychometric scale development; the responsibility questionnaire's internal consistency is treated as the primary reliability indicator because the affective claim is the one under scrutiny. Table 4 places the cognitive and affective outcomes side by side to make the asymmetry immediately visible.

Although the study foregrounds the descriptive asymmetry, inferential tests are reported for completeness. For the cognitive test the gain was statistically reliable and moderate in magnitude: the mean difference of 9.29 points (95% CI [5.81, 12.78]) was significant on a paired-samples t-test, $t(16) = 5.65$, $p < .001$, and on the Wilcoxon signed-rank test, $W = 0$, $Z = -3.64$, $p < .001$, with a between-condition effect of Cohen's $d = 0.63$ (pooled SD). For the responsibility self-report, by contrast, the mean difference of 2.00 points (95% CI [-0.40, 4.40]) was not statistically significant, $t(16) = 1.76$, $p = .097$ (Wilcoxon $W = 36$, $Z = -1.66$, $p = .097$; $d = 0.39$). Because the confidence interval for the affective change includes zero, the observed two-point shift is statistically indistinguishable from no change, which supports rather than contradicts the descriptive reading.

Table 4. Summary of the Cognitive–Affective Asymmetry in the Field Trial

Dimension	Outcome in the field trial
Cognitive (knowing)	Normalised gain $g = 0.37$ (medium), ceiling-bounded; pretest mean 74.59, posttest mean 83.88.
Responsibility (self-report)	81% to 83% (+2 points), within measurement noise; no category transition; both measurements in the very high band.

Discussion

A Cognitive–Affective Gap

The headline result of this study is the asymmetry between the cognitive and the affective outcome, and the interpretation of that asymmetry given here differs in tone from the way comparable results are usually reported. A moderate, ceiling-bounded cognitive gain of 0.37 in normalised gain coexists with an affective shift of two percentage points that is best read as measurement noise.

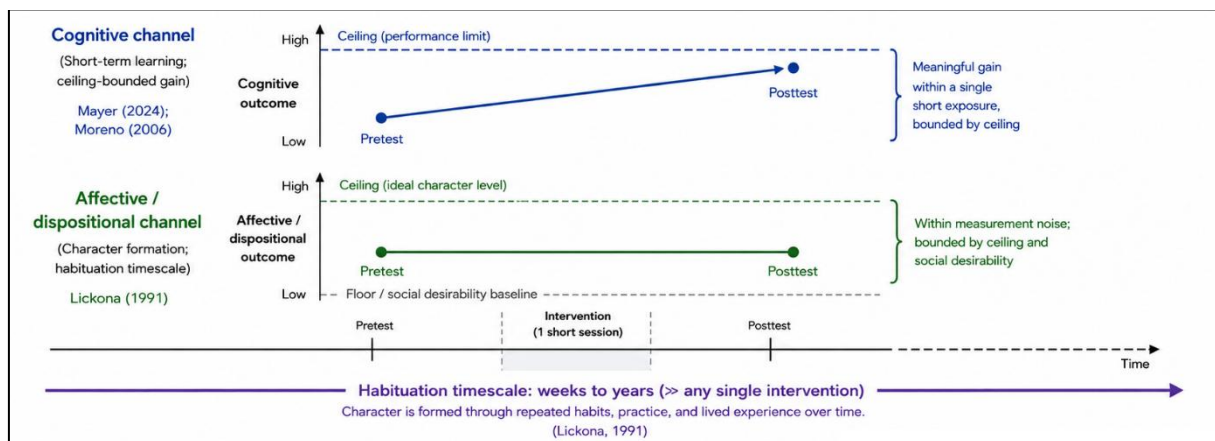


Figure 4. Proposed Interpretive Framework for the Cognitive–Affective Asymmetry

Three Sources of Affective Optimism

Three structural features of the prevailing template combine to produce a systematic bias toward affective optimism, and each is supported by an established methodological literature. The first is the use of the same instrument at pretest and posttest without a control group, which [Spurlock \(2018\)](#) and [Bierer et al. \(2025\)](#) have argued cannot in principle separate intervention effects from history, maturation, and testing effects. The second is the routine use of student self-report on character constructs that are already culturally desirable. [Camerini & Schulz \(2018\)](#) and [Ng et al. \(2025\)](#) have shown that primary-school children systematically inflate self-report scores on socially valued constructs, with younger children showing higher levels of this inflation than older students. Indonesian fifth-grade students are likely to enter a responsibility measurement at scores already above 80 per cent, which leaves almost no numerical room for the medium to demonstrate further effect, and to exit the measurement at scores still above 80 per cent regardless of any actual change in their behaviour outside the questionnaire. The third is the convention of interpreting any positive shift as evidence of effectiveness, including shifts of one or two

percentage points that fall within the noise band of the instrument. Each feature is defensible in isolation. In combination, they may produce an evaluation pipeline in which a positive affective conclusion appears likely to be reached, regardless of whether real character change has occurred. This pattern is consistent with the broader publication-bias dynamics documented by [Ropovik et al. \(2021\)](#), [Evans & Yuan \(2022\)](#), and [Kraft \(2023\)](#) in educational research more generally.

Implications for the Wider Evidence Base

This critique applies to the present study in the first instance. The present study used the prevailing template, and the affective result of 81 per cent moving to 83 per cent is precisely the kind of small, ceiling-compressed, noise-bounded shift that the template makes such outcomes structurally likely. The difference is that the shift is read here as measurement noise rather than as evidence of effectiveness. The question then arises, by extension, what proportion of the affective claims in the second cluster of the literature might dissolve under the same reading. This study does not claim that all such studies overstate their findings. The more cautious claim is that the evidentiary standard those studies meet is the same standard the present study meets, and that this standard does not support strong claims about character change. [Brown et al. \(2023\)](#) reached an analogous conclusion in their meta-analysis, observing that selection bias in published character-education studies systematically inflated effect estimates, with corrected effects smaller than the published averages would suggest.

One feature of the meta-analytic evidence requires direct acknowledgement, because it appears to cut against the present argument. [Brown et al. \(2023\)](#) report that single-session interventions were associated with larger, not smaller, effects than longer programmes. Read naively, this might seem to license exactly the short-intervention optimism criticised here. Two considerations temper that reading. First, [Brown et al.](#) note that the number of single-session interventions in their corpus was small, so the estimate is correspondingly unstable. Second, and more importantly, their effect sizes derive from controlled comparisons against a control condition, whereas the inference contested in this study is the attribution of dispositional change on the basis of an uncontrolled, single-group, self-report design. The present argument is therefore not that brevity in itself precludes effect, but that brevity combined with a dispositional outcome and a ceiling-prone self-report instrument cannot support the affective claims routinely made under the prevailing template.

A Mismatch of Timescales

The theoretical reading of the asymmetry leads to a similar conclusion. [Mayer \(2024\)](#) explicitly limits the dependent variables of the Cognitive Theory of Multimedia Learning to retention and transfer, and [Noetel et al. \(2021\)](#) confirmed that video produces reliable cognitive gains in higher education with effect sizes that are nevertheless small to moderate when video replaces existing teaching. [Moreno \(2006\)](#) extended the framework with the affective mediation hypothesis, but the affective constructs that mediate cognition in CATLM are situational interest, motivation, and emotional valence, all of which are short-lived states rather than stabilised dispositions. Neither theory was designed to explain long-term dispositional change. The finding reported here therefore does not falsify multimedia learning theory. It exposes a level-of-analysis gap. Multimedia theory operates at the level of single learning episodes lasting minutes to hours, while character formation operates at the level of habituation lasting weeks to years across multiple social contexts ([Arthur et al., 2016](#); [Walker & Thoma, 2017](#)). To bridge this gap, the field would need a theory of multimedia in habituation rather than a theory of multimedia in single-session learning. CATLM is a step in the right direction, but its explanatory targets remain too narrow for the dispositional outcomes that character education claims to produce.

These considerations can be consolidated into a simple two-level framework for evaluating media in character education, which clarifies what each kind of outcome requires. At the first level, affect is treated as a process variable in the sense of CATLM: situational interest, satisfaction, or confusion that arises within a single learning episode, is legitimately measured by state self-report immediately after exposure, and is the appropriate target for a short animated video. At the second level, character is treated as a disposition: a stabilised pattern of feeling and action that develops through habituation across weeks and contexts, that cannot be validly captured by an immediate post-exposure self-report, and that requires multi-session interventions, behavioural or observational indicators, delayed measurement, and a control condition. The recurrent error in the literature is to use first-level instruments and designs to license second-level claims. The framework predicts the asymmetry reported here: a single video can move a process variable detectably but leaves a dispositional measure essentially unchanged, and any apparent

movement on the latter is more plausibly read as measurement fluctuation than as character change. Figure 4 summarises this distinction.

Implications for Design and Practice

The contrast with the first cluster of studies clarifies what stronger evidence looks like. [Widiyasanti & Ayryza \(2018\)](#) reported a significant t-test of 3.810 on responsibility character, but their design used separate experimental and control schools and a longer treatment, both of which subtract part of the testing and ceiling biases that affect single-class designs. [Wuryanti & Kartowagiran \(2016\)](#) reached similar findings on the value of hard work using comparable design choices. The difference between these studies and the second cluster is not the medium and not the topic. It is the design strength, the dose, and the instrument. The implication is direct. A field that wishes to make defensible claims about character formation through media must invest in design strength, dose, and instrument quality, not in further variations of the medium itself.

This implication has practical weight. Much of the published development literature on character-oriented animated video in Indonesian elementary research has been organised around the prevailing template, including theses and postgraduate research. If a meaningful portion of the affective effectiveness reported under that template is artefactual, then the same effort directed instead at teacher development, family-school programmes, and structured habituation activities, in line with the multi-component approach advocated by [Brown et al. \(2023\)](#) and [Arthur et al. \(2016\)](#), might yield larger affective returns. This study does not claim a definitive verdict on this question. It claims, instead, that the question is now visible and worth asking, and that visible questions are the precondition for redirecting effort.

For classroom teachers, the practical recommendation that follows is specific. Animated video of the kind developed here is best positioned not as a standalone character intervention but as the knowledge-activation and framing component that opens a longer habituation sequence ([Hastuti et al., 2026](#)). A defensible sequence would use the video to establish shared concepts and vocabulary, then move to teacher modelling, guided practice, structured reflection, and repeated opportunities to enact the target behaviours, with reinforcement extended into the family and the wider school community. On this understanding, the moderate cognitive gain documented here is precisely the contribution the medium is suited to make, while the work of forming disposition is carried by the surrounding programme rather than by the video alone.

Concretely, teachers can treat a single animated video as the opening move of a multi-week sequence rather than a self-contained lesson. The video can introduce shared vocabulary and a vivid anchoring example; subsequent sessions can add teacher modelling and think-alouds of the target behaviour; recurring, structured opportunities can let students practise that behaviour in authentic classroom and playground situations with specific feedback; short home tasks can extend it beyond school; and growth can be monitored through observation and teacher and parent report across weeks rather than a single post-lesson questionnaire. For curriculum developers, the corresponding recommendation is to specify ([Rakhmawati et al., 2023](#)), alongside each animated product, the surrounding programme of practice and the longitudinal, multi-source assessment against which its contribution to character should be judged.

None of this implies that one-group pretest–posttest designs have no place. For the early, formative stages of instructional-product development—feasibility checks, usability refinement, and piloting before a product is ready for rigorous testing—such designs are efficient and appropriate, and the present study used one for precisely that developmental purpose. The cognitive benefits of well-designed multimedia are, by contrast, comparatively well established ([Mayer, 2024](#)); the difficulty arises only when a formative, feasibility-oriented design is asked to license summative claims about dispositional change. Once effectiveness for character formation becomes the question, the design must change accordingly: multisite, controlled, and longitudinal evaluations, such as the cluster-randomised trial of an elementary character programme reported by [Zhang et al. \(2025\)](#), illustrate the level of evidence that dispositional claims require.

Alternative Explanations Considered

Two objections to this argument deserve direct response. The first is that the sample of seventeen students at one school is too small to support claims about the broader literature. The sample is indeed small, and no inferential claim is made from it about other studies. The claim is therefore interpretive and methodological rather than inferential. The same methodological features that bound the affective result reported here will bound studies employing the same methodological configuration, regardless of sample

size. Larger samples within the same template do not remove the testing, ceiling, and social desirability effects. They only estimate them with greater precision (Kraft, 2023). The second objection is that the reading offered here, which treats the affective result as noise, risks dismissing real, if small, character gains. This risk is accepted as the more honest of two errors. Treating noise as signal accumulates spurious findings across the literature, while treating small signal as noise accumulates conservative null results. The first error is harder to correct than the second.

CONCLUSION

This study developed an animated video on rights, obligations, and responsibility for fifth-grade Civic Education and tested it within the dominant evaluation template used in Indonesian character-education media research. The product passed validation, the cognitive gain was moderate but ceiling-bounded, and the affective shift was within measurement noise. Read together, these results support a single thesis. Within the conditions of this study, animated video supported knowing more clearly than becoming. The cognitive channel through which the Cognitive Theory of Multimedia Learning explains learning operates well within a single short exposure. The dispositional channel through which character is actually formed does not, and asking a single short video intervention to produce stable dispositional change may exceed what the instrument and intervention are designed to support.

The contribution of this study is therefore not the product, which serves primarily as a vehicle for examining the methodological question, and not the cognitive gain, which is ceiling-bounded. The principal contribution is methodological rather than product-oriented. Development research on media for character education will produce more credible knowledge if it abandons the assumption that affective effectiveness can be inferred from a one-group pretest-posttest comparison on student self-report. Three priorities follow. First, character-oriented media should be evaluated as components of longer instructional sequences across weeks rather than as standalone products tested within a single session, ideally with control conditions and multi-school, longitudinal designs that permit causal inference. Second, affective evaluation should be triangulated through teacher observation, peer report, and parent report, with delayed posttests that allow social desirability to attenuate. Third, the field should consider whether a research programme that has produced many feasible products and few defensible claims about character change is using its resources well, or whether some portion of those resources should be redirected toward the broader social and pedagogical conditions on which character formation actually depends.

These contributions must be read against the study's limitations, consolidated here. The evidence comes from a single intact class of seventeen fifth-grade students in one public elementary school, delivered in a single session; the small sample, single-site setting, and short duration together mean that the empirical results are not statistically generalisable and should be treated as illustrative rather than representative. What transfers beyond this setting is not the numerical result but the methodological point—the asymmetry that the prevailing template produces—which does not depend on sample size.

REFERENCES

- Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for Learning: Methods and Development* (3rd ed.). Boston: Allyn and Bacon.
- Arthur, J., Kristjánsson, K., Harrison, T., Sanderse, W., & Wright, D. (2016). *Teaching character and virtue in schools*. (1st ed.) (Citizenship, Character and Values Education). Routledge. <https://doi.org/10.4324/9781315695013>
- Bierer, S. B., Beck Dallaghan, G., Borges, N. J., Brondfield, S., Fung, C. C., Huggett, K. N., Teal, C. R., Thammasitboon, S., & Colbert, C. Y. (2025). Moving Beyond Simplistic Research Design in Health Professions Education: What a One-Group Pretest-Posttest Design Will Not Prove. *MedEdPORTAL*. https://doi.org/10.15766/mep_2374-8265.11527
- Brown, M., McGrath, R. E., Bier, M. C., Johnson, K., & Berkowitz, M. W. (2023). A comprehensive meta-analysis of character education programs. *Journal of Moral Education*, 52(2), 119–138. <https://doi.org/10.1080/03057240.2022.2060196>
- Camerini, A.-L., & Schulz, P. J. (2018). Social Desirability Bias in Child-Report Social Well-Being: Evaluation of the Children's Social Desirability Short Scale Using Item Response Theory and Examination of Its Impact on Self-Report Family and Peer Relationships. *Child Indicators Research*, 11(4), 1159–1174. <https://doi.org/10.1007/s12187-017-9472-9>

- Evans, D. K., & Yuan, F. (2022). How Big Are Effect Sizes in International Education Studies? *Educational Evaluation and Policy Analysis*, 44(3), 532–540. <https://doi.org/10.3102/01623737221079646>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hastuti, W. S., Rakhmawati, Y., Firdaus, H. F. M., & Condro, R. (2026). Engagement : Game-Based STEM Learning for Enhanced Engagement : A Collaborative Program Involving Teachers , Students , and Parents. *Engagement: Jurnal Pengabdian Kepada Masyarakat*, 10(01), 51–64.
- Kleftodimos, A. (2024). Computer-Animated Videos in Education: A Comprehensive Review and Teacher Experiences from Animation Creation. *Digital*, 4(3), 613–647. <https://doi.org/10.3390/digital4030031>
- Kraft, M. A. (2023). The Effect-Size Benchmark That Matters Most: Education Interventions Often Fail. *Educational Researcher*, 52(3), 183–187. <https://doi.org/10.3102/0013189X231155154>
- Lickona, T. (1991). *Educating for Character: How Our Schools Can Teach Respect and Responsibility*.
- Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition*, 10(2), 229–240. <https://doi.org/10.1016/j.jarmac.2021.03.007>
- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8. <https://doi.org/10.1007/s10648-023-09842-1>
- Mayer, R. E., & Fiorella, L. (Eds.). (2021). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press. <https://doi.org/10.1017/9781108894333>
- Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, 22(3), 149–158. <https://doi.org/10.1111/j.1365-2729.2006.00170.x>
- Munjiatun, M. (2018). Penguatan Pendidikan Karakter: Antara Paradigma dan Pendekatan. *Jurnal Kependidikan*, 6(2), 334–349. <https://doi.org/10.24090/jk.v6i2.1924>
- Ng, Z. J., Lin, S., Niu, L., & Cipriano, C. (2025). Measurement Invariance of the Children's Social Desirability Scale–Short Version (CSD-S) Across Gender, Grade Level, and Race/Ethnicity. *Assessment*, 32(3), 394–404. <https://doi.org/10.1177/10731911241245789>
- Nissen, J. M., Talbot, R. M., Nasim Thompson, A., & Van Dusen, B. (2018). Comparison of normalized gain and Cohen's *d* for analyzing gains on concept inventories. *Physical Review Physics Education Research*, 14(1), 010115. <https://doi.org/10.1103/PhysRevPhysEduRes.14.010115>
- Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video Improves Learning in Higher Education: A Systematic Review. *Review of Educational Research*, 91(2), 204–236. <https://doi.org/10.3102/0034654321990713>
- Oldham, P., & McLoughlin, S. (2026). Character education empirical research: A thematic review and comparative content analysis. *Journal of Moral Education*, 55(2), 313–341. <https://doi.org/10.1080/03057240.2025.2480185>
- Purwati, P., Japar, M., Qomariyah, L., & Tentama, F. (2024). Moral knowing, moral feeling, and moral action in reflecting moral development of students in junior high school. *International Journal of Evaluation and Research in Education (IJERE)*, 13(3), 1602. <https://doi.org/10.11591/ijere.v13i3.25499>
- Putri, S. A., Rohmani, Apriza, B., & Elizar. (2024). Effectiveness of Using Animation Videos in Science Learning in Elementary Schools: A Systematic Literature Reviews. *Indonesian Journal of Educational Research and Review*, 7(3), 667–678. <https://doi.org/10.23887/ijerr.v7i3.82242>
- Rakhmawati, Y., Elitasari, H. T., Yulianti, I., Safitri, B., & Asip, M. (2023). Relevansi Kurikulum 2013 dan Efikasi Diri Dalam Mempersiapkan Tuntutan Pendidikan Abad 21. *Khazanah Pendidikan-Jurnal Ilmiah Kependidikan*, 17(1), 34–42. <https://doi.org/10.30595/jkp.v17i1.15092>
- Ropovik, I., Adamkovic, M., & Greger, D. (2021). Neglect of publication bias compromises meta-analyses of educational research. *PLOS ONE*, 16(6), e0252415. <https://doi.org/10.1371/journal.pone.0252415>
- Spurlock, D. R. (2018). The Single-Group, Pre- and Posttest Design in Nursing Education Research: It's Time to Move on. *Journal of Nursing Education*, 57(2), 69–71. <https://doi.org/10.3928/01484834-20180123-02>
- Walker, D. I., & Thoma, S. J. (2017). Moral and Character Education. In *Oxford Research Encyclopedia of Education*. <https://doi.org/10.1093/acrefore/9780190264093.013.119>

- Widiyasanti, M., & Ayriza, Y. (2018). Pengembangan Media Video Animasi Untuk Meningkatkan Motivasi Belajar Dan Karakter Tanggung Jawab Siswa Kelas V. *Jurnal Pendidikan Karakter*, 9(1). <https://doi.org/10.21831/jpk.v8i1.21489>
- Wuryanti, U., & Kartowagiran, B. (2016). Pengembangan Media Video Animasi Untuk Meningkatkan Motivasi Belajar Dan Karakter Kerja Keras Siswa Sekolah Dasar. *Jurnal Pendidikan Karakter*, 7(2). <https://doi.org/10.21831/jpk.v6i2.12055>
- Zhang, Y., Larson, M., & Cook, C. R. (2025). A Multi-Site Cluster Randomized Controlled Trial Evaluating the Effect of *CharacterStrong's PurposeFull People* Elementary Program on Student Outcomes. *Canadian Journal of School Psychology*, 40(2), 79–101. <https://doi.org/10.1177/08295735251340750>