SCHOULID : Indonesian Journal of School Counseling (2026), 11(2), 16-33
ISSN (Print): 2548-3234 | ISSN (Electronic): 2548-3226
DOI: 10.23916/086868011

ELORA Open Access Journal: https://schoulid.eloracenter.org/schoulid



Effectiveness of psychoeducation-based guided reflective journaling in improving students' mental health literacy: a single-group pretest-posttest study

Khairul Bariyyah*)[1], Raissa Dwifandra Putri[1], Zihan Novita Sari[1], Ayun Ulfatul Jannah[2], Siti Nurhaliza[1], Parid Rilo Pambudi[3], Santy Andrianie[4], Laily Tiarani Soejanto[5]

Universitas Negeri Malang[1], SMP Laboratorium Universitas Negeri Malang[2], Universitas Islam Balitar[3], Universitas Sebelas Maret[4], Universitas PGRI Kanjuruhan Malang[5]

*) Correspondence regarding this article should be addressed to: Author address e-mail: khairul.bariyyah.fip@um.ac.id

Abstract: Mental Health Literacy (MHL) is essential for early problem recognition and timely help-seeking among adolescents. However, evidence regarding school-based MHL programs in Indonesia remains limited. This preliminary study examined changes in MHL following participation in Jurnal SEJIWA, a psychoeducation-based guided journaling program integrated into school counseling. Thirty junior high school students (17 girls and 13 boys; mean age = 13.53 years) from Malang, Indonesia, completed the Universal Mental Health Literacy Scale for Adolescents (UMHL-A) before and after the intervention. Significant improvements were observed across all four domains: help-seeking, stigma-related attitudes, mental health knowledge, and knowledge of mental illnesses (all Holm-adjusted $p < .001$). All participants demonstrated gains in help-seeking and stigma-related attitudes, with no score decreases across the assessed domains. However, internal consistency was poor ($\alpha = -.233$ to $.249$), raising concerns about score reliability. Exploratory analyses revealed no significant differences by sex, grade, or age after multiple-comparison adjustment. Despite substantial pre–post improvements, the single-group design limits causal interpretation. Further research using controlled designs and culturally validated instruments is needed to establish the effectiveness of Jurnal SEJIWA.

Keywords: Mental health literacy, Guided reflective journaling, Psychoeducation, School counseling, Adolescents

Article History: Received on 15/08/2026; Revised on 26/10/2026; Accepted on 27/10/2026; Published Online: 13/11/2026.

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2026 by author.

Introduction

Adolescence is a time when many begin to encounter mental health challenges. A summary analyzing 192 studies from various populations estimated that roughly one out of three individuals with any psychiatric condition start showing signs before age 14, and almost half before they turn 18, with the highest incidence near age 14.5 (Solmi et al., 2022). In Indonesia, analyses of nationally representative adolescent mental health surveys indicate substantial unmet needs: only 2.6% of adolescents with a diagnostic or subthreshold mental disorder accessed mental health or counseling

services in the preceding year (Wahdi et al., 2025). These results highlight the crucial role school counseling plays in early identification and prevention for Indonesian adolescents.

The term "mental health literacy" (MHL) originally described people's information and attitudes regarding mental illness that help them spot, deal with, or avoid such problems (Cresswell-Smith et al., 2026). The focus has since expanded to highlight practical knowledge that enables individuals to act for their own wellness and support others (Soria-Martínez et al., 2024). Kutcher et al., (2016) introduced a four-component model that includes: knowing ways to sustain and improve one's mental health; awareness of mental disorders and treatment options; actions or attitudes that help counteract stigma; and beliefs or skills that enhance the intention and ability to seek help. Such a model is especially suited for schools since it connects a strengths-based knowledge of mental health with illness awareness and attitudes affecting support-seeking. Despite these developments, how MHL is measured still varies widely. Recent reviews likewise document substantial variation in mental health literacy measures and gaps in evidence about instruments used to recognize mental health difficulties in children (Johnson et al., 2023). The Universal Mental Health Literacy Scale for Adolescents (UMHL-A) was recently developed to cover all four of Kutcher's elements efficiently for youth ages 10–14 (Kågström et al., 2023).

MHL's main impact for teens tends to be through seeking support. A more recent systematic review identified stigma, practical barriers, negative attitudes toward services, and gaps in mental health knowledge as recurring obstacles to adolescent help-seeking (Barrow & Thomas, 2022). A more recent review of 53 studies found that nearly all obstacles stemmed from youth's own characteristics, such as not knowing enough about mental health, and most also mentioned social hurdles like the fear of stigma (Radez et al., 2021). Studies in countries with fewer resources also describe youth not recognizing signs of mental health issues, widespread negative stereotypes, and hesitation or doubt about professional help (Renwick et al., 2024). Education can tackle these obstacles, so schools increasingly host MHL programs.

Research on school-based MHL interventions shows mostly encouraging results, but not all evidence is consistent. In a cluster-randomized trial with 534 Canadian high schoolers across 24 schools, a curriculum taught by teachers raised knowledge and lowered stigma compared to controls (Milin et al., 2016). A review of randomized studies found several school programs improved MHL and reduced stigma, but relatively few followed up with participants later (Ma, Anderson, et al., 2023). A companion analysis found limited data backing effects on help-seeking and linked this issue to the shortage of standardized and validated measures in that domain (Ma, Burn, et al., 2023). One meta-analysis pooling studies with follow-up (on average 5 months after programs) reported ongoing progress in MHL ($d = 0.48$) and more modest effects on negative attitudes ($d = 0.30$) and unwillingness to interact with peers facing difficulties ($d = 0.16$; Freţian et al., 2021). For children and teens, interventions aiming to boost help-seeking have led to short-term improvements in knowledge and reduction in personal stigma, but have not consistently shifted actual use of professional supports (Xu et al., 2018). In the Indonesian context, the IMPeTUs initiative partnered with local collaborators to create a digital, culturally tailored MHL program for ages 11–15, following a review that showed few available options fit local needs (Brooks et al., 2021), and further case studies indicated that students and educators found it both acceptable and feasible (Brooks et al., 2023). However, there is little published data on school-based, counselor-led MHL activities in Indonesia. Related Indonesian school-counseling initiatives have trained counselors in cognitive-behavioral stress management and peer-support approaches to address students' academic stress (Bariyyah et al., 2024). These initiatives provide contextual support for counselor-led programs but do not directly establish the effects of guided reflective journaling on mental health literacy.

Many such school programs emphasize direct teaching: providing information through lessons or handouts on mental health and mental illness. But relaying facts does not necessarily shape

attitudes like willingness to talk about emotional struggles or continue friendships with classmates who are struggling, because these changes often rely on how students connect new knowledge with their own lives (Nobre et al., 2021). Activities built around writing introduce a structured method to foster this personal reflection. Still, it is key to recognize the differences in writing types. Expressive writing invites participants to engage with personally meaningful emotional experiences, although the instructions and topics can differ substantially between programs (Hu et al., 2025). A recent meta-analysis of 31 controlled studies with follow-up assessments found small improvements in depression, anxiety, and stress symptoms following expressive writing ($g = -0.12$; Guo (2023). A separate school-based trial reported sustained improvement in anxiety when structured narrative writing was paired with a brief cognitive-behavioral program (Y. Wang et al., 2026). Journaling focused on gratitude encourages reflection on good experiences; in a randomized study of 70 Indonesian junior high school students, a structured gratitude-journaling program improved positive affect relative to a control condition (Amalia et al., 2026). A synthesis of 20 randomized trials testing journaling strategies for people experiencing mental illness found mild to moderate benefits, but noted results varied greatly and most studies had methodological limitations (Sohal et al., 2022).

Guided reflective journaling takes a different approach. Unlike open-ended writing or gratitude exercises, and distinct from stand-alone psychoeducational sessions, it provides students with concise educational content and then prompts them to relate this material to their daily life through structured questions—for example, by asking them to identify an emotion felt during the previous week, its cause, and possible ways to seek help. This difference is important, as evidence shows the nature of writing tasks can shift the outcomes. A recent systematic review of educational journaling found that guided formats were common and could facilitate reflection, although effects differed across contexts and assessment methods (Brown et al., 2022). An experimental study further found that feedback on reflective journal entries could improve subsequent reflective quality, supporting the use of structured guidance rather than writing tasks alone (Pieper et al., 2021).

There are several theoretical pathways through which structured reflection might influence MHL, though these were not directly examined in the current study. First, psychoeducational materials offer tools and frameworks—such as how thoughts, emotions, and actions relate—that students might use to make sense of what happens in their lives. Second, writing about one's own experiences using these frameworks could support meaning-making, although successful meaning-making depends on how experiences are interpreted and connected with the self (Huang et al., 2021). Third, the act of giving words to feelings—affect labeling—may influence emotional intensity and the regulation of feelings (Givon et al., 2026), potentially lowering barriers to talking about emotions and seeking support. Fourth, reflecting on how ordinary individuals encounter stress and feelings might help reduce social distance from peers managing mental health struggles, which can influence stigma. These are presented as possible mechanisms guiding the intervention's design, not as outcomes measured in this research.

Jurnal SEJIWA (Self-Help Jiwa Sehat untuk Siswa, "self-help for healthy minds for students") is a guided reflection and journaling program combining educational materials on mental health topics with structured writing, delivered as part of regular school counseling services. This study reports an initial, single-group analysis of change in MHL knowledge, attitudes, and help-seeking among junior high schoolers who completed Jurnal SEJIWA activities. Because the research did not include a control group, it intended to summarize the direction, scale, and pattern of change, not to claim causality. The study addressed five main questions: (1) How do students' UMHL-A domain results shift from baseline to final assessment? (2) How large, precise, and consistent within individuals are these observed changes? (3) Are there descriptive differences in initial scores or subsequent gains for boys versus girls or by grade? (4) Does age relate to variations in change across domains? (5) What

design features and measurement shortcomings limit the interpretation? Questions 3 and 4 were exploratory and not part of the initial preregistration.

METHOD

The study used a quantitative single-group pretest-posttest design. All participants completed the UMHL-A before and after taking part in Jurnal SEJIWA activities, and scores were paired using unique student codes. The design does not include a comparison group and therefore cannot isolate the intervention from maturation, repeated testing, facilitator attention, school events, or other concurrent influences (Cuijpers et al., 2017). The intervention description was organized with reference to the child- and adolescent-focused adaptation of the Template for Intervention Description and Replication (TIDieR-C) (Goren et al., 2025).

The study took place at SMP Laboratorium Universitas Negeri Malang, a junior high school in Malang, East Java, Indonesia. Thirty students who were participating in the school's guidance and counseling services took part. Sex was recorded in the dataset as jenis kelamin (sex) and is reported as sex; gender identity was not measured. Of the 30 participants, 13 (43.3%) were male and 17 (56.7%) were female; 16 (53.3%) were in Grade VII and 14 (46.7%) were in Grade VIII. Ages ranged from 12 to 15 years (M = 13.53, SD = 1.07). Table 1 presents participant characteristics, including the distribution of age within each grade.

Jurnal SEJIWA is a structured reflective journal designed to help students understand mental health through reflective writing and simple psychological exercises. According to the study materials, its content covers (a) an introduction to mental health and mental health awareness, (b) identification of emotions and awareness of one's own psychological state, (c) recognition of sources of stress, (d) relationships among thoughts, feelings, and behavior, (e) strategies for managing emotions, and (f) the importance of seeking psychological help. Each activity combines psychoeducational material with trigger questions, reflection guides, and written exercises intended to help students connect the material with their everyday experiences.

School counselors (guru Bimbingan dan Konseling) facilitated the activities. Their documented roles were to explain the material, guide the reflective process, help students understand their psychological experiences, and reinforce students' understanding of mental health. The activities were designed so that students could explore their experiences with structured support without being required to disclose personal information that made them uncomfortable. Written reflections and counselor observation notes were collected during implementation; these qualitative materials were not available for this report and were not analyzed.

MHL was measured with the UMHL-A (Kågström et al., 2023), a 17-item self-report scale that covers the four components of Kutcher's framework. Part A contains eight attitude items rated from 1 (strongly disagree) to 5 (strongly agree), with an additional "I don't know" option. Part B contains nine knowledge statements answered "Yes," "No," or "I don't know." Following published UMHL-A scoring conventions described in validation research (Uysal et al., 2025), and consistent with Clark et al., (2025), items were assigned to four domains: help-seeking (HS; A1, A2, A4, A6, A7; possible range 5-25), stigma-related attitudes (ST; A3, A5, A8; 3-15), knowledge of mental health (KMH; B1, B6, B7, B8, B9; 0-5), and knowledge of mental illnesses (KMI; B2, B3, B4, B5; 0-4). In Part A, "I don't know" is scored 3. In Part B, correct ("Yes") responses are scored 1 and "No" and "I don't know" responses are scored 0. The ST items are worded so that higher scores indicate more favorable, less stigmatizing attitudes; they were not reverse-scored. The workbook supplied for analysis contained already-scored binary Part B values, so incorrect and "I don't know" responses could not be separated. Because the four domains use different response formats and have not been shown to

form a single interval-level composite, domains were analyzed separately and no total score was computed.

The UMHL-A was developed through a literature review, a Delphi study, focus groups, and cognitive interviews with adolescents, and its structure was refined in a representative sample of Czech adolescents (Kågström et al., 2023). Validation studies in other countries support a two-part, four-factor structure. In Turkey, the scale was adapted for adolescents aged 10-14 years and judged valid and reliable (Ciydem & Avci, 2024). A Chinese version yielded Cronbach's alpha coefficients of .80 and .78 for the attitude and knowledge subscales (Q. Wang et al., 2024), and an Australian validation reported McDonald's ω of .78 for HS, .67 for ST, .72 for KMH, and .68 for KMI (Clark et al., 2025). Internal consistency was estimated in the present sample and is reported in the Results.

The study proceeded in three phases. In the first phase, all 30 participants completed the UMHL-A as a pretest. In the second phase, students took part in Jurnal SEJIWA activities within the guidance and counseling program, facilitated by school counselors. In the third phase, after the intervention was completed, students completed the same UMHL-A as a posttest.

Pretest and posttest worksheets were merged by student code, and variables were selected by column name rather than position because the posttest worksheet included a demographic column absent from an earlier version. All 30 codes were matched one-to-one; sex, grade, and age were identical across occasions; and all 17 item responses were present and within their permissible ranges at both occasions. The analysis used the observations as recorded, without modification. The number of independent units was 30 students; the 510 paired item comparisons (30 students × 17 items) were used only descriptively.

For each domain and occasion we report the mean, standard deviation, median, interquartile range, and observed minimum and maximum. Change was defined as posttest minus pretest. Distributions of paired differences were inspected using frequency tables and Shapiro-Wilk tests. Because the domains are short sums of discrete items, change scores take few distinct values and contain many ties; normality tests and parametric assumptions should therefore be interpreted cautiously.

Following the supplied analysis plan, two-sided Wilcoxon signed-rank tests were the primary tests because the change distributions were inconsistent with normality. W is defined as the sum of ranks of the positive differences (posttest > pretest), with absolute differences ranked using average ranks for ties. Zero differences were excluded before ranking (Wilcoxon's method), and p values were obtained from the normal approximation with tie correction and continuity correction, because exact p values are not available in the software when ties or zeros are present. The Wilcoxon test was interpreted as a test of whether positive and negative differences are equally likely to occur with equal rank magnitudes, not as a direct test of the mean difference. The Holm step-down procedure (Guilbaud, 2022) controlled the familywise error rate across the four domain tests.

For each domain we report the mean change with a t-based 95% confidence interval (CI); the numbers of students whose scores increased, did not change, or decreased; and the matched-pairs rank-biserial correlation, $r_{rb} = (W+ - W-)/(W+ + W-)$, which equals the proportion of favorable minus unfavorable rank evidence (Ben-Shachar et al., 2020). As an additional standardized measure of within-participant change, we report Cohen's d_z, the mean change divided by the standard deviation of change (Dankel & Loenneke, 2021). Its 95% CI was obtained from a participant-level paired percentile bootstrap with 20,000 resamples and random seed 20261110. Because d_z is scaled by the variability of change rather than by between-person variability, it is not comparable with between-group effect sizes and does not indicate a controlled treatment effect (Cuijpers et al., 2017). Paired-samples t tests were run as sensitivity analyses for mean change. Because symmetry of the

differences could not be assumed, two-sided exact sign tests and Wilcoxon tests with Pratt's treatment of zeros were also run as sensitivity analyses.

Demographic analyses were added to complement the main analysis and are explicitly exploratory. For sex and grade, we report subgroup pretest, posttest, and change scores and compare change scores directly between groups using Welch's independent-samples t test, which does not assume equal variances (Delacre et al., 2017). Differences are expressed as female minus male and Grade VIII minus Grade VII. Because several subgroup change distributions departed from normality, Mann-Whitney tests were run as sensitivity analyses. The Mann-Whitney test evaluates whether a change score drawn from one group tends to exceed one drawn from the other; it is not, in general, a test of medians (Divine et al., 2018). For age, which took only four values, we computed Spearman's rank correlation between age and change in each domain, using average ranks for ties, a two-sided permutation p value (20,000 permutations), and a percentile bootstrap 95% CI (20,000 resamples). Age was kept continuous and was not dichotomized.

The four sex comparisons, four grade comparisons, and four age associations were treated as one declared exploratory family of 12 tests. False discovery rate was controlled with the Benjamini-Hochberg procedure (García-Pérez, 2023) using the Welch p values for sex and grade and the permutation p values for age. This family was kept separate from the four main Holm-adjusted tests. Age and grade overlapped substantially: all six 12-year-olds were in Grade VII, all seven 15-year-olds were in Grade VIII, and the Spearman correlation between age and grade was .73. Age and grade findings were therefore not interpreted as independent. As a further sensitivity analysis, we fitted parsimonious ordinary least squares models in which the posttest domain score was predicted by the pretest score and one demographic variable at a time, with conventional and HC3 robust standard errors and Shapiro-Wilk checks of residuals. No multivariable model with several demographic variables or interactions was fitted, and no sex × age × grade subgroups were examined.

Internal consistency was estimated separately for each domain at each occasion using Cronbach's alpha (Taber, 2018)), which for the binary Part B items is equivalent to KR-20 (Madadizadeh & Bahariniya, 2025). We report percentile bootstrap 95% CIs (20,000 resamples), mean inter-item correlations, and the range of corrected item-total correlations. Alpha was interpreted as an index of the average covariation among items, not as evidence of unidimensionality (McNeish, 2018). No items were removed, and confirmatory factor analysis, structural equation modeling, and longitudinal invariance testing were not attempted with 30 participants. We also report item-level change and the number of students at the maximum possible score to examine ceiling effects.

Analyses were conducted in Python 3.13.18 with NumPy 2.5.3 (Harris et al., 2020), pandas 3.0.5, SciPy 1.18.1 (Virtanen et al., 2020), and statsmodels 0.15.0; figures were produced with Matplotlib. All random procedures used the seed 20261110. The full analysis script, which reproduces every reported value from the workbook, is available from the corresponding author on reasonable request. Preliminary descriptive values and alpha estimates supplied before analysis were reproduced exactly.

Results and Discussion

All 30 students had paired pretest and posttest records with complete responses on all 17 items (Table 1). The sex distribution was similar across grades (Grade VII: 6 male, 10 female; Grade VIII: 7 male, 7 female), but age and grade were strongly related. Grade VII comprised students aged 12-14 years, and Grade VIII comprised students aged 13-15 years.

Table 1. Participant Characteristics

Characteristic	Total (*N* = 30)	Grade VII (*n* = 16)	Grade VIII (*n* = 14)
Sex			
Male	13 (43.3)	6	7
Female	17 (56.7)	10	7
Grade			
Grade VII	16 (53.3)	16	—
Grade VIII	14 (46.7)	—	14
Age (years)			
12	6 (20.0)	6	0
13	9 (30.0)	7	2
14	8 (26.7)	3	5
15	7 (23.3)	0	7
M (*SD*)	13.53 (1.07)	—	—
Range	12-15	12-14	13-15

Note. Values in the Total column are *n* (%) unless otherwise indicated; grade columns show counts. Sex was recorded as *jenis kelamin* (sex); gender identity was not measured. Age and grade overlapped substantially (Spearman ϱ = .73): all 12-year-olds were in Grade VII and all 15-year-olds were in Grade VIII.

Table 1 summarizes scores at each occasion and the within-participant changes; Figure 1 shows individual trajectories and occasion-specific means.

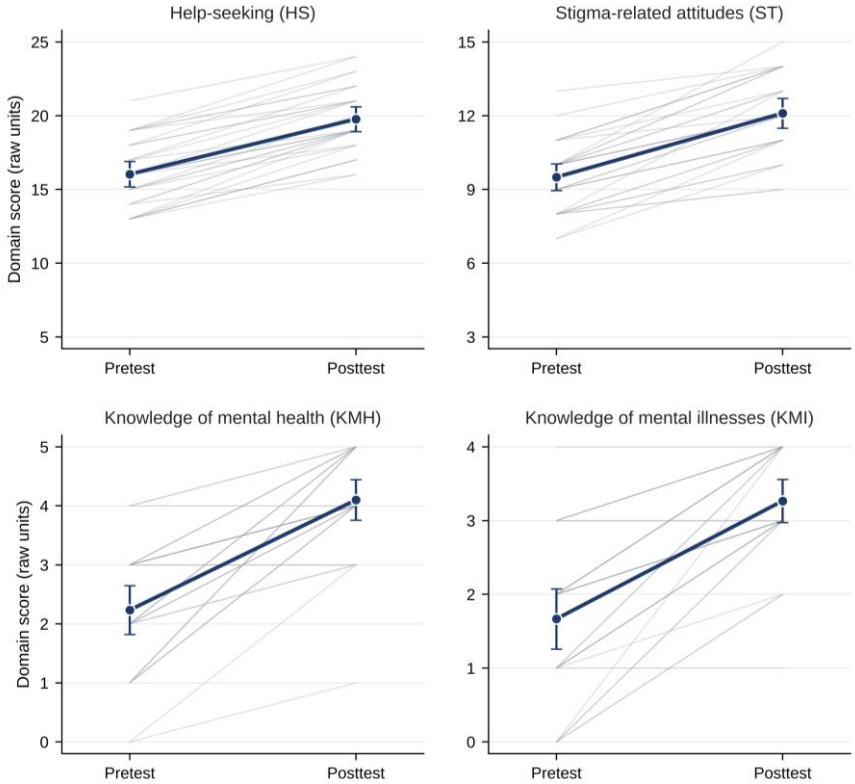


Figure 1. Individual and Mean Changes Across Mental Health Literacy Domains

Note. Each thin gray line represents one student (*N* = 30); lines are semi-transparent, and because scores are discrete many trajectories overlap and appear as a single darker line. The dark blue line connects occasion-specific means, and error bars are 95% confidence intervals. Panels use original domain units and different *y*-

axis scales spanning each domain's possible range (HS 5-25, ST 3-15, KMH 0-5, KMI 0-4), so panel heights are not comparable across domains.

Mean scores were higher at posttest than at pretest in every domain. HS increased from 16.03 (*SD* = 2.31) to 19.77 (*SD* = 2.24), a mean change of 3.73 points (95% CI [3.32, 4.15]) on a 5-25 scale. ST increased from 9.50 (*SD* = 1.46) to 12.10 (*SD* = 1.63), a mean change of 2.60 points (95% CI [2.20, 3.00]) on a 3-15 scale. KMH increased from 2.23 (*SD* = 1.10) to 4.10 (*SD* = 0.92), a mean change of 1.87 points (95% CI [1.42, 2.31]) on a 0-5 scale, and KMI increased from 1.67 (*SD* = 1.09) to 3.27 (*SD* = 0.78), a mean change of 1.60 points (95% CI [1.23, 1.97]) on a 0-4 scale. These values reproduced the preliminary reference values exactly.

Table 2. Descriptive Scores and Within-Participant Changes Across UMHL-A Domains

Panel A. Descriptive statistics by occasion

Domain	Range	Pretest			Posttest		
		M (SD)	Mdn [Q1, Q3]	Min-max	M (SD)	Mdn [Q1, Q3]	Min-max
Help-seeking (HS)	5-25	16.03 (2.31)	16.00 [14.00, 18.00]	13-21	19.77 (2.24)	19.00 [18.25,21.00]	16-24
Stigma-related attitudes (ST)	3-15	9.50 (1.46)	10.00 [8.00, 10.00]	7-13	12.10 (1.63)	12.00 [11.00, 13.75]	9-15
Knowledge of mental health (KMH)	0-5	2.23 (1.10)	2.00 [1.25, 3.00]	0-4	4.10 (0.92)	4.00 [4.00, 5.00]	1-5
Knowledge of mental illnesses (KMI)	0-4	1.67 (1.09)	2.00 [1.00, 2.00]	0-4	3.27 (0.78)	3.00 [3.00, 4.00]	1-4

Panel B. Within-participant change (posttest − pretest)

Domain	Mean change [95% CI]	SD_{change}	$n\uparrow$ / $n=$ / $n\downarrow$	W ($n_{\neq 0}$)	p	p_{Holm}	r_{rb}	d_z [95% CI]
Help-seeking (HS)	3.73 [3.32, 4.15]	1.11	30 / 0 / 0	465 (30)	< .001	< .001	1.00	3.36 [2.83, 4.30]
Stigma-related attitudes (ST)	2.60 [2.20, 3.00]	1.07	30 / 0 / 0	465 (30)	< .001	< .001	1.00	2.43 [2.00, 3.29]
Knowledge of mental health (KMH)	1.87 [1.42, 2.31]	1.20	27 / 3 / 0	378 (27)	< .001	< .001	1.00	1.56 [1.23, 2.07]
Knowledge of mental illnesses (KMI)	1.60 [1.23, 1.97]	1.00	26 / 4 / 0	351 (26)	< .001	< .001	1.00	1.59 [1.23, 2.21]

Note. N = 30 students. UMHL-A = Universal Mental Health Literacy Scale for Adolescents. Higher scores indicate more favorable help-seeking attitudes, more favorable (less stigmatizing) attitudes, and greater knowledge. Mean change CIs are t-based. $n\uparrow$ / $n=$ / $n\downarrow$ = number of students whose domain score increased, did not change, or decreased. W = sum of ranks of positive differences in the two-sided Wilcoxon signed-rank test; zero differences were excluded and p values use the normal approximation with tie and continuity corrections. p_{Holm} = Holm-adjusted across the four domain tests. r_{rb} = matched-pairs rank-biserial correlation. d_z = mean change divided by the SD of change; CIs are percentile bootstrap intervals (20,000 participant-level resamples, seed 20261110). d_z describes within-participant change and is not a controlled treatment effect. Domains are in different raw units and should not be compared on raw scores.

Change was consistent across individuals. All 30 students had higher HS and ST scores at posttest. For KMH, 27 students improved and 3 were unchanged; for KMI, 26 improved and 4 were unchanged. No student's score decreased in any domain. At the item level, 274 of the 510 paired item comparisons (53.7%) showed an increase, 236 (46.3%) showed no change, and none showed a decrease. This absence of any decrease is an observed feature of the dataset and is considered further in the Discussion.

Change scores were discrete, with five or six distinct values per domain, and Shapiro-Wilk tests indicated departures from normality in all four domains (W = .898-.910, p = .008-.015). Skewness was small to moderate (0.09-0.54). Because every nonzero difference was positive, the Wilcoxon statistic reached its maximum possible value in each domain: W = 465 (n = 30 nonzero differences) for HS and ST, W = 378 (n = 27) for KMH, and W = 351 (n = 26) for KMI. The corresponding normal approximations were z = 4.81, 4.82, 4.58, and 4.52, all p < .001 before and after Holm adjustment, and the rank-biserial correlation was 1.00 in each domain. Exact conditional p values would be smaller still, because a result in which all nonzero differences share the same sign has a two-sided probability of $2/2^n$ under the null hypothesis. Given that the differences were exclusively positive, the Wilcoxon results are best read as evidence of a consistent upward shift; a symmetric location-shift interpretation adds little beyond this.

Standardized within-participant change was large: d_z = 3.36 (95% bootstrap CI [2.83, 4.30]) for HS, 2.43 [2.00, 3.29] for ST, 1.56 [1.23, 2.07] for KMH, and 1.59 [1.23, 2.21] for KMI. The very large HS and ST values reflect small variability in change (*SD*s of change = 1.11 and 1.07) as much as the size of the mean change. Sensitivity analyses gave the same conclusions: paired t tests, exact sign tests, and Wilcoxon tests with Pratt's treatment of zeros were all significant at p < .001.

Knowledge scores approached the ceiling at posttest. Eleven students (36.7%) obtained the maximum KMH score and 13 (43.3%) obtained the maximum KMI score, compared with none and one, respectively, at pretest. No student reached the maximum HS score, and one reached the maximum ST score at posttest.

Internal consistency was very low at both occasions. Alpha coefficients were .048 and .073 for HS, −.013 and −.233 for ST, .021 and .169 for KMH, and .249 and .005 for KMI at pretest and posttest, respectively. Bootstrap CIs were wide and, in every case, included zero; mean inter-item correlations ranged from −.060 to .080, and corrected item-total correlations were often near zero or negative. These estimates indicate that, in this sample, items within each domain covaried very little, so the summed domain scores cannot be assumed to reflect coherent underlying attributes. The small sample increases the uncertainty of these estimates but does not by itself explain values this close to or below zero. The domain scores are therefore interpreted below as sums of the specified items, and changes in them are interpreted as changes in item responses rather than as established changes in latent constructs. Table 3 and Figure 2 present changes by sex and by grade. Both male and female students and both grade groups showed higher mean scores at posttest in every domain.

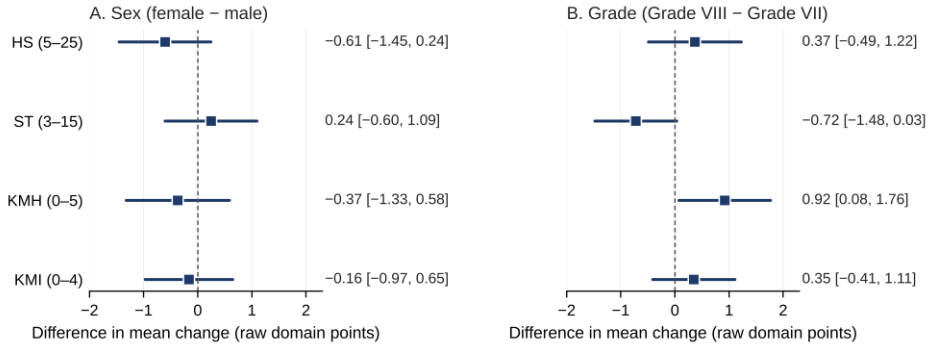


Figure 2. Exploratory Mean Changes by Sex and Grade

Note. Squares show the difference in mean change between groups, and horizontal lines show Welch 95% confidence intervals; values are printed to the right. Panel A subtracts males from females (n = 17 vs. 13); Panel B subtracts Grade VII from Grade VIII (n = 14 vs. 16). Positive values indicate larger increases in the female or Grade VIII group. Differences are in each domain's own raw units (ranges in parentheses), so magnitudes are not directly comparable across rows. No difference was significant after Benjamini-Hochberg correction. Exploratory analysis

Differences in mean change between groups were small and imprecisely estimated. For sex (female minus male), the differences were −0.61 points for HS (95% CI [−1.45, 0.24]), 0.24 for ST [−0.60, 1.09], −0.37 for KMH [−1.33, 0.58], and −0.16 for KMI [−0.97, 0.65]; no raw or adjusted *p* value fell below .05. For grade (Grade VIII minus Grade VII), the differences were 0.37 points for HS [−0.49, 1.22], −0.72 for ST [−1.48, 0.03], 0.92 for KMH [0.08, 1.76], and 0.35 for KMI [−0.41, 1.11]. The KMH grade difference had an unadjusted *p* of .034, but its Benjamini-Hochberg-adjusted *p* was .168. Grade VII students began with somewhat higher KMH scores than Grade VIII students (2.44 vs. 2.00), and after adjustment for pretest KMH, the estimated grade difference was smaller and its CI included zero (*b* = 0.62, 95% CI [−0.03, 1.27]). Mann-Whitney sensitivity tests led to the same pattern of conclusions.

Table 3. Exploratory Changes by Sex and Grade

Panel A. Sex

	Male (*n* = 13)			Female (*n* = 17)			Female − male			
Domain	Pre *M* (*SD*)	Post *M* (*SD*)	Change [95% CI]	Pre *M* (*SD*)	Post *M* (*SD*)	Change [95% CI]	Δ change [95% CI]	*t* (*df*)	*p*	p_{BH}
HS	15.23 (2.01)	19.31 (1.84)	4.08 [3.36, 4.79]	16.65 (2.40)	20.12 (2.50)	3.47 [2.95, 3.99]	−0.61 [−1.45, 0.24]	−1.48 (23.5)	.153	.366
ST	9.62 (1.71)	12.08 (1.85)	2.46 [1.74, 3.19]	9.41 (1.28)	12.12 (1.50)	2.71 [2.20, 3.21]	0.24 [−0.60, 1.09]	0.60 (23.0)	.556	.607
KMH	2.15 (1.14)	4.23 (0.60)	2.08 [1.24, 2.91]	2.29 (1.10)	4.00 (1.12)	1.71 [1.17, 2.24]	−0.37 [−1.33, 0.58]	−0.81 (21.7)	.428	.514
KMI	1.85 (1.28)	3.54 (0.52)	1.69 [0.98, 2.41]	1.53 (0.94)	3.06 (0.90)	1.53 [1.08, 1.98]	−0.16 [−0.97, 0.65]	−0.42 (21.3)	.681	.681

Panel B. Grade

	Grade VII (*n* = 16)			Grade VIII (*n* = 14)			Grade VIII − Grade VII			
Domain	Pre *M* (*SD*)	Post *M* (*SD*)	Change [95% CI]	Pre *M* (*SD*)	Post *M* (*SD*)	Change [95% CI]	Δ change [95% CI]	*t* (*df*)	*p*	p_{BH}
HS	16.69 (2.41)	20.25 (2.44)	3.56 [3.05, 4.08]	15.29 (2.02)	19.21 (1.93)	3.93 [3.20, 4.66]	0.37 [−0.49, 1.22]	0.88 (24.1)	.388	.514
ST	9.50 (1.55)	12.44 (1.46)	2.94 [2.34, 3.54]	9.50 (1.40)	11.71 (1.77)	2.21 [1.70, 2.73]	−0.72 [−1.48, 0.03]	−1.96 (27.8)	.060	.179
KMH	2.44 (1.09)	3.88 (1.02)	1.44 [0.85, 2.02]	2.00 (1.11)	4.36 (0.74)	2.36 [1.69, 3.02]	0.92 [0.08, 1.76]	2.23 (27.0)	.034	.168
KMI	1.75 (1.29)	3.19 (0.75)	1.44 [0.92, 1.95]	1.57 (0.85)	3.36 (0.84)	1.79 [1.18, 2.39]	0.35 [−0.41, 1.11]	0.94 (26.7)	.355	.514

Note. Exploratory analyses. Change = posttest − pretest; subgroup change CIs are *t*-based. Δ change = difference in mean change between groups, with the reference group subtracted as labeled (female − male; Grade VIII − Grade VII), estimated with Welch's *t* test and its 95% CI. p_{BH} = Benjamini-Hochberg-adjusted *p* within the declared family of 12 exploratory tests (Table 4, Panel B). HS = help-seeking (range 5-25); ST = stigma-related attitudes (3-15); KMH = knowledge of mental health (0-5); KMI = knowledge of mental illnesses (0-4). A difference in change between groups should not be interpreted as treatment moderation.

Table 4 and Figure 3 show the associations between age and change. Age was positively associated with change in KMH (ρ_s = .46, 95% CI [.13, .73], permutation *p* = .012) and negatively associated with change in ST (ρ_s = −.37, 95% CI [−.66, −.02], *p* = .042). Associations with change in HS (ρ_s = .18) and KMI (ρ_s = .16) were weak and imprecise. After Benjamini-Hochberg correction across the 12 exploratory tests, no association remained below .05 (adjusted *p* = .142 for KMH and .168 for ST). In baseline-adjusted sensitivity models, older age was associated with higher posttest KMH (*b* = 0.42 per year, 95% CI [0.13, 0.70]) and slightly lower posttest ST (*b* = −0.361, 95% CI [−0.720, −0.002]);

these models were not adjusted for multiplicity. Because age and grade were strongly related (ϱ_s = .73), the age and grade patterns for KMH are not independent findings.

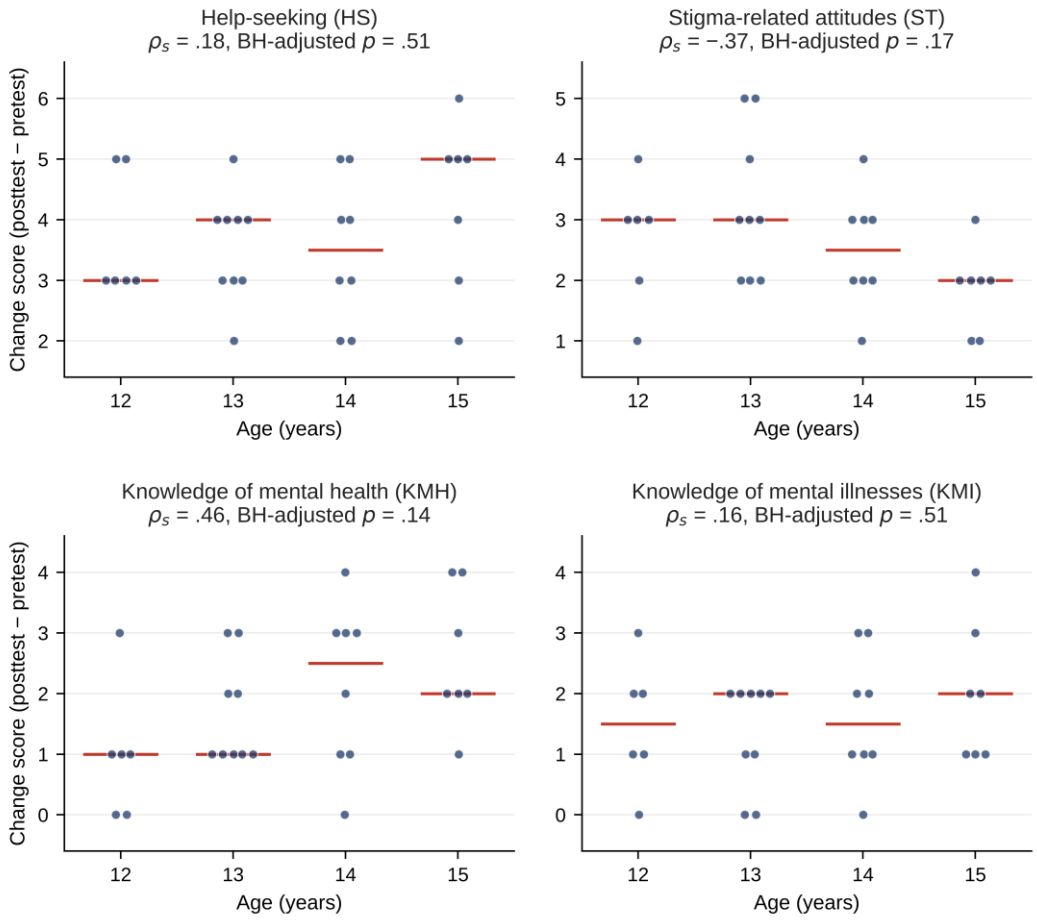


Figure 3. Age and Changes in Mental Health Literacy

Note. Exploratory analysis. Each point is one student; points with the same age and change score are offset horizontally for display only, and all analyses used original ages. Red horizontal bars mark the median change at each age. ϱ_s = Spearman correlation; BH-adjusted p = Benjamini-Hochberg-adjusted permutation p value within the family of 12 exploratory tests. No fitted trend lines are shown. Age and grade overlap substantially, so these associations are not independent of grade.

Table 4. Exploratory Associations Between Age and Domain Changes

Panel A. Spearman correlations between age and change scores

Domain change	**ϱ_s**	**95% CI**	**p_{perm}**	**p_{BH}**
Help-seeking (HS)	.18	[−.22, .55]	.352	.514
Stigma-related attitudes (ST)	−.37	[−.66, −.02]	.042	.168
Knowledge of mental health (KMH)	.46	[.13, .73]	.012	.142
Knowledge of mental illnesses (KMI)	.16	[−.22, .50]	.389	.514

Panel B. Complete family of 12 exploratory tests

Comparison	Domain	Test statistic	Raw p	p_{BH}
Sex (female − male)	HS	Welch $t(23.5) = -1.48$	.153	.366
Sex (female − male)	ST	Welch $t(23.0) = 0.60$	.556	.607
Sex (female − male)	KMH	Welch $t(21.7) = -0.81$	.428	.514
Sex (female − male)	KMI	Welch $t(21.3) = -0.42$	.681	.681
Grade (VIII − VII)	HS	Welch $t(24.1) = 0.88$	.388	.514
Grade (VIII − VII)	ST	Welch $t(27.8) = -1.96$	.060	.179
Grade (VIII − VII)	KMH	Welch $t(27.0) = 2.23$	.034	.168
Grade (VIII − VII)	KMI	Welch $t(26.7) = 0.94$	.355	.514
Age (continuous)	HS	Spearman $\varrho_s = .18$	.352	.514
Age (continuous)	ST	Spearman $\varrho_s = -.37$	.042	.168
Age (continuous)	KMH	Spearman $\varrho_s = .46$	.012	.142
Age (continuous)	KMI	Spearman $\varrho_s = .16$	.389	.514

Note. $N = 30$. Age was analyzed as a continuous variable with four observed values (12-15 years); Spearman correlations use average ranks for ties. p_{perm} = two-sided permutation p value (20,000 permutations); 95% CIs are percentile bootstrap intervals (20,000 resamples). p_{BH} = Benjamini-Hochberg false discovery rate adjustment across all 12 tests in Panel B; this family is separate from the four Holm-adjusted main tests. Age and grade were strongly associated ($\varrho_s = .73$), so age and grade results are not independent.

This initial school-based evaluation found that junior high school students' scores on all four UMHL-A domains were higher after participation in Jurnal SEJIWA, a counselor-facilitated, psychoeducation-based guided reflective journaling program. The increases were consistent in direction across students and statistically robust across primary and sensitivity analyses. At the same time, two features of the evidence limit what can be concluded. The design had no comparison group, so the increases cannot be attributed to the intervention, and the domain scores showed very weak internal consistency, so it is unclear what attributes the summed scores represent. We therefore interpret the findings as observed improvement following the intervention, not as established effectiveness.

The largest change in raw units was in help-seeking, where the mean rose by nearly four points on a 5-25 scale. Item-level results show increases in all five HS items, including statements about whether talking about feelings helps and willingness to talk with peers and adults. Because "I don't know" is scored at the scale midpoint, part of this change could reflect students moving from uncertainty to a definite view, which is itself a plausible consequence of psychoeducation. Stigma-related attitudes also improved for every student. The ST domain is positively keyed, so higher scores indicate more favorable attitudes, such as willingness to remain friends with a peer who develops a mental health problem. These attitudinal shifts are consistent with the intended content of the journal, which asked students to reflect on emotions, stress, and help-seeking in their own lives, and with the theoretical view that personalizing information may make attitudes more receptive to change (Huang et al., 2021).

Knowledge of mental health and of mental illnesses also increased, with the largest item-level gains on statements that relatively few students endorsed at pretest. For example, the proportion endorsing the statement that how people get along with others affects how they feel rose from 23% to 73%, and the proportion recognizing depression as one of the most common mental illnesses among young people rose from 37% to 87%. Because many students reached the maximum score at posttest, the knowledge domains may underestimate the range of change among the most knowledgeable students and are less able to discriminate among them. Because Part B responses

were supplied as scored values, we could not distinguish students who moved from an incorrect answer to a correct one from those who moved from "I don't know" to a correct answer; recording raw response options would permit this distinction in future studies.

The magnitude of change was larger than typically reported in controlled school-based MHL studies. For example, a meta-analysis of interventions with follow-up assessments reported a pooled MHL effect of d = 0.48 relative to controls (Freţian et al., 2021), and school-based RCTs have generally shown modest knowledge and stigma effects (Ma, Anderson, et al., 2023; Milin et al., 2016). These values are not directly comparable with ours. The dz statistic divides mean change by the standard deviation of change, and in a single-group design it also absorbs testing effects, maturation, and expectancy (Cuijpers et al., 2017; Dankel & Loenneke, 2021). The very small variability of change in HS and ST, rather than an unusually large raw change, is the main reason those dz values exceeded 2. The raw changes and their CIs are therefore the more informative summary of what was observed.

The internal consistency estimates were well below those reported in validation studies of the UMHL-A in other countries Clark et al., (2025) Wang et al., 2024) and below conventional thresholds at both occasions. Several explanations are possible and cannot be distinguished here. Short domains with three to five items yield unstable alpha estimates in small samples, and binary knowledge items with restricted variance at posttest further attenuate covariation. Items may also have been understood differently in the Indonesian administration, whose translation and adaptation procedures still need to be documented. The ST domain illustrates a content issue: its items concern the importance of getting along with others, recommending professional help to someone with a long-term problem, and remaining friends with a peer who has a mental health problem, which are related to, but not obviously indicators of, a single stigma attitude. Alpha reflects average inter-item covariation and does not by itself establish or refute unidimensionality (McNeish, 2018), but values near or below zero do mean that the summed scores should not be treated as precise measures of coherent constructs in this sample.

The combination of incoherent domain structure with uniformly positive change is notable. Items within a domain barely covaried with one another at either occasion, yet 274 of 510 item responses increased and none decreased. In short self-report scales, some item-level decreases between occasions are usually expected because of ordinary response variability. The absence of any decrease could reflect a genuine and consistent learning effect across all content. It could also be influenced by features of the setting, such as students responding in ways they perceived as expected after a counselor-led program, familiarity with the items after the pretest, close correspondence between intervention content and questionnaire statements, or the conditions under which the posttest was administered. The data were confirmed by the authors as original records, and this observation is not evidence of error. However, documenting the administration conditions, retaining the original response forms, and having data entry independently verified would strengthen confidence in the pattern. Future evaluations could reduce expectancy effects by having questionnaires administered by staff not involved in the intervention and by assuring students that responses do not affect their standing with counselors.

No sex, grade, or age difference in change survived false discovery rate correction, and the estimated differences between groups were small relative to their uncertainty. The data therefore do not provide evidence that the program's observed changes differed by sex, grade, or age. Two unadjusted signals deserve brief comment because they could inform future hypotheses. Older students and Grade VIII students showed larger KMH gains, but Grade VII students had higher KMH scores at pretest, so part of this pattern may reflect baseline differences and other biases affecting uncontrolled comparisons, including regression toward the mean (Cham et al., 2024). Older students also showed slightly smaller ST gains. With 30 participants, overlapping age and grade groups, and 12 exploratory tests, these patterns should be regarded as hypothesis-generating at

most. They should not be interpreted as evidence of treatment moderation, which would require a controlled design and adequate subgroup sizes.

With the above limitations in mind, the findings suggest that a guided reflective journal can be integrated into routine school counseling services in Indonesia and that students' MHL scores improved in the period during which they used it. This is relevant in a context in which only a small proportion of Indonesian adolescents with mental health needs access formal services (Wahdi et al., 2025). Journaling resources are inexpensive and can be delivered by school counselors, and structured prompts may help students move from describing events to reflecting on emotions, stress, and sources of support (Brown et al., 2022; Pieper et al., 2021). Counselors who adopt such resources should protect the privacy of written reflections, establish clear procedures for responding to distress or risk disclosed in journals, and avoid making disclosure a requirement. It should also be emphasized that higher literacy scores are not the same as actual help-seeking, symptom reduction, or improved mental health. Reviews have found little evidence that school MHL programs change help-seeking behavior (Ma, Burn, et al., 2023; Xu et al., 2018), so programs should be evaluated on these outcomes directly rather than assuming them. A systematic review of psychoeducational group leadership also emphasizes the role of counselor facilitation and purposeful guidance in group-based services (Nuha et al., 2024).

Several limitations qualify these findings. First, the single-group design cannot rule out alternative explanations for the observed change, including repeated testing, maturation over the intervention period, facilitator attention, expectancy and social desirability, and concurrent school activities (Cuijpers et al., 2017). Second, the measurement evidence was weak: internal consistency was very low or negative in every domain at both occasions, the Indonesian adaptation of the UMHL-A has not been documented or validated, and Part B scores could not distinguish incorrect from "I don't know" responses. Third, knowledge scores approached the ceiling at posttest. Fourth, the sample was small and drawn from one laboratory school associated with a university, which may differ from other Indonesian junior high schools, and recruitment procedures have not yet been documented. Fifth, subgroup analyses were based on small groups, age and grade were confounded, and multiple tests were performed. Sixth, there was no follow-up assessment, so the durability of change is unknown. Seventh, key implementation information, including dosage, attendance, and fidelity, was not available. Finally, qualitative data from journals and counselor observations were collected but not analyzed, so the processes through which students engaged with the intervention could not be examined.

A stronger evaluation would use a controlled design, such as a cluster-randomized or stepped-wedge trial across classes or schools, with a preregistered protocol and an active or usual-practice comparison condition. The Indonesian UMHL-A should be adapted and validated systematically, including cognitive interviews with Indonesian adolescents, factor analysis in an adequately sized sample, and tests of measurement invariance over time before change scores are compared across occasions. Raw response options, including "I don't know," should be recorded. Outcomes should extend beyond literacy to help-seeking intentions and behavior, school counseling contacts, and well-being, with follow-up assessments months after the program. Implementation data on attendance, journal completion, and fidelity, together with systematic qualitative analysis of de-identified reflections, would clarify how students engage with guided reflection and which components contribute to change.

CONCLUSIONS

This study provides preliminary evidence of improvements in adolescents' mental health literacy following participation in Jurnal SEJIWA, a counselor-facilitated psychoeducational guided

journaling program. Significant improvements were observed across all four UMHL-A domains: help-seeking, stigma-related attitudes, knowledge of mental health, and knowledge of mental illnesses. These findings highlight the potential of guided reflective journaling as an accessible and student-centered approach to strengthening mental health literacy within school counseling services. However, the absence of a comparison group and the low internal consistency of the UMHL-A domains limit causal interpretation and the reliability of the observed changes. Therefore, the findings should be interpreted as favorable changes following program participation rather than evidence of established intervention effectiveness. Future research should employ controlled designs, culturally adapted and psychometrically validated instruments, larger and more diverse samples, and follow-up assessments. Further evaluation should also consider actual help-seeking behaviors, students' reflective experiences, and program implementation outcomes. Overall, Jurnal SEJIWA offers an initial empirical foundation for integrating guided reflective journaling into school-based mental health literacy education and preventive counseling services.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Universitas Negeri Malang, Indonesia, for funding this research through the 2026 Research Collaboration Scheme of the Pendidikan Profesi Guru (PPG) and SMP Laboratorium Universitas Negeri Malang. This research was financially supported by the Pendidikan Profesi Guru (PPG) Collaboration Fund under Research Implementation Agreement No. 14.4.2037/UN32.14.1/LT/2026, dated April 14, 2026, and Rector's Decree No. 13.4.281/UN32/KP/2026, dated April 13, 2026. The authors also extend their appreciation to SMP Laboratorium Universitas Negeri Malang, particularly the school counselors, participating students, and school administrators, for their cooperation and support throughout the implementation of the Jurnal SEJIWA program.

REFERENCES

Amalia, K. E., Pradipta, P. K., Aldilla, D. N., Srisayekti, W., Rubiyanti, Y., & Purnama, C. Y. (2026). Gratitude Journaling Practices Influences Positive Affect among Early Adolescents: A Study among Junior High School Students in a Rural Area Rancaekek Indonesia. *G-Couns: Jurnal Bimbingan Dan Konseling*, *10*(03 SE-Articles), 2764–2777. https://doi.org/10.31316/g-couns.v10i03.8319

Bariyyah, K., Atmoko, A., Zen, E. F., Apriani, R., & Maizura, N. (2024). Optimalisasi Peran Guru Bimbingan dan Konseling melalui Cognitive Behavior Stress Management untuk Mendukung Peer Counselor dalam Penanggulangan Stres Akademik Siswa. *Abdimas Pedagogi: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, *7*(2), 116–125. https://doi.org/10.17977/um050v7i22024p116-125

Barrow, E., & Thomas, G. (2022). Exploring perceived barriers and facilitators to mental health help-seeking in adolescents: a systematic literature review. *Educational Psychology in Practice*, *38*(2), 173–193. https://doi.org/10.1080/02667363.2022.2051441

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815)

Brooks, H., Irmansyah, I., Syarif, A. K., Pedley, R., Renwick, L., Rahayu, A. P., Manik, C., Prawira, B., Hann, M., Brierley, H., Lovell, K., & Bee, P. (2023). Evaluating a prototype digital mental health literacy intervention for children and young people aged 11–15 in Java, Indonesia: a mixed methods, multi-site case study evaluation. *Child and Adolescent Psychiatry and Mental Health*, *17*(1), 79. https://doi.org/10.1186/s13034-023-00608-9

Brooks, H., Syarif, A. K., Pedley, R., Irmansyah, I., Prawira, B., Lovell, K., Opitasari, C., Ardisasmita, A., Tanjung, I. S., Renwick, L., Salim, S., & Bee, P. (2021). Improving mental health literacy among young people aged 11–15 years in Java, Indonesia: the co-development of a culturally-

appropriate, user-centred resource (The IMPeTUs Intervention). *Child and Adolescent Psychiatry and Mental Health*, *15*(1), 56. https://doi.org/10.1186/s13034-021-00410-5

Brown, L. G., Briscoe, G. S., & Grabowsky, A. (2022). The Influence of Journaling on Nursing Students: A Systematic Review. *Journal of Nursing Education*, *61*(1), 29–35. https://doi.org/10.3928/01484834-20211203-01

Cham, H., Lee, H., & Migunov, I. (2024). Quasi-experimental designs for causal inference: an overview. *Asia Pacific Education Review*, *25*(3), 611–627. https://doi.org/10.1007/s12564-024-09981-2

Ciydem, E., & Avci, D. (2024). Psychometric properties of the Turkish version of the universal mental health literacy scale for adolescents. *Journal of Pediatric Nursing: Nursing Care of Children and Families*, *79*, e186–e191. https://doi.org/10.1016/j.pedn.2024.10.020

Clark, J. C., Bartik, W., & Rice, K. (2025). Australian validation of the Universal Mental Health Literacy Scale for Adolescents. *Australian Journal of Psychology*, *77*(1), 2555651. https://doi.org/10.1080/00049530.2025.2555651

Cresswell-Smith, Johanna, Sultson, Hedvig, Murd, Carolina, Havik, Merle, Griebler, Robert, Schütze, Denise, Bevilacqua, Roberta, Sgolastra, Arianna, Laidra, Kaia, Lárusdóttir, Sólrún Ósk, Árnadóttir, Ösp, & Solin, Pia. (2026). Conceptualisation and operationalisation of mental health literacy: An umbrella review. *Scandinavian Journal of Public Health*, 14034948261422936. https://doi.org/10.1177/14034948261422936

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, *26*(4), 364–368. https://doi.org/10.1017/S2045796016000809

Dankel, S. J., & Loenneke, J. P. (2021). Effect Sizes for Paired Data Should Use the Change Score Variability Rather Than the Pre-test Variability. *Journal of Strength & Conditioning Research*, *35*(6), 1773–1778. https://doi.org/10.1519/JSC.0000000000002946

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology*, *30*(1). https://doi.org/10.5334/irsp.82

Divine, G. W., Norton, H. J., Barón, A. E., & Juarez-Colunga, E. (2018). The Wilcoxon–Mann–Whitney Procedure Fails as a Test of Medians. *The American Statistician*, *72*(3), 278–286. https://doi.org/10.1080/00031305.2017.1305291

Freţian, A. M., Graf, P., Kirchhoff, S., Glinphratum, G., Bollweg, T. M., Sauzet, O., & Bauer, U. (2021). The Long-Term Effectiveness of Interventions Addressing Mental Health Literacy and Stigma of Mental Illness in Children and Adolescents: Systematic Review and Meta-Analysis. *International Journal of Public Health*, *Volume 66*. https://doi.org/10.3389/ijph.2021.1604072

García-Pérez, M. A. (2023). Use and misuse of corrections for multiple testing. *Methods in Psychology*, *8*, 100120. https://doi.org/10.1016/j.metip.2023.100120

Givon, E., Meiran, N., & Goldenberg, A. (2026). The process of affect labeling. *Trends in Cognitive Sciences*, *30*(5), 422–433. https://doi.org/10.1016/j.tics.2025.09.017

Goren, K., Lai, V. K. W., Baba, A., Smith, M., Allegaert, K., Butcher, N. J., Fernandes, R., Gill, P. J., Hartling, L., Juszczak, E., Longmuir, P. E., Macarthur, C., Mehta, K., Metz, D., Nguyen, K. A., Odermarsky, M., Poluru, R., Purper-Ouakil, D., Reggiardo, G., … Offringa, M. (2025). Transparent Reporting of Pediatric Clinical Trial Interventions: TIDieR-Children and Adolescents. *Pediatrics*, *156*(5), e2025072867. https://doi.org/10.1542/peds.2025-072867

Guilbaud, O. J. M. (2022). Note on a stepwise procedure for rejecting at least k out of m hypotheses: A simple Holm-type formulation and proof. *Biometrical Journal*, *64*(2), 377–383. https://doi.org/10.1002/bimj.202000206

Guo, L. (2023). The delayed, durable effect of expressive writing on depression, anxiety and stress: A meta-analytic review of studies with long-term follow-ups. *British Journal of Clinical Psychology*, *62*(1), 272–297. https://doi.org/10.1111/bjc.12408

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hu, Chao S, Zhang, Haotian, Hu, Lihao, Nadiiev, Artur, Yan, Junru, & Liu, Mengyuan. (2025). Efficacy of positive expressive writing in reducing depression and social anxiety symptoms among schoolchildren during COVID-19: A randomized controlled trial. *Journal of Health Psychology*, *30*(10), 2516–2528. https://doi.org/10.1177/13591053241295446

Huang, M., Schmiedek, F., & Habermas, T. (2021). Only some attempts at meaning making are successful: The role of change-relatedness and positive implications for the self. *Journal of Personality*, *89*(2), 175–187. https://doi.org/10.1111/jopy.12573

Johnson, C. L., Gross, M. A., Jorm, A. F., & Hart, L. M. (2023). Mental Health Literacy for Supporting Children: A Systematic Review of Teacher and Parent/Carer Knowledge and Recognition of Mental Health Problems in Childhood. *Clinical Child and Family Psychology Review*, *26*(2), 569–591. https://doi.org/10.1007/s10567-023-00426-7

Kågström, A., Pešout, O., Kučera, M., Juríková, L., & Winkler, P. (2023). Development and validation of a universal mental health literacy scale for adolescents (UMHL-A). *Psychiatry Research*, *320*, 115031. https://doi.org/10.1016/j.psychres.2022.115031

Kutcher, Stan, Wei, Yifeng, & Coniglio, Connie. (2016). Mental Health Literacy: Past, Present, and Future. *The Canadian Journal of Psychiatry*, *61*(3), 154–158. https://doi.org/10.1177/0706743715616609

Ma, K. K. Y., Anderson, J. K., & Burn, A.-M. (2023). Review: School-based interventions to improve mental health literacy and reduce mental health stigma – a systematic review. *Child and Adolescent Mental Health*, *28*(2), 230–240. https://doi.org/10.1111/camh.12543

Ma, K. K. Y., Burn, A.-M., & Anderson, J. K. (2023). Review: School-based mental health literacy interventions to promote help-seeking – a systematic review. *Child and Adolescent Mental Health*, *28*(3), 408–424. https://doi.org/10.1111/camh.12609

Madadizadeh, F., & Bahariniya, S. (2025). Tutorial on internal consistency assessment by Cronbach's alpha and McDonald's omega. *Perioperative Care and Operating Room Management*, *41*, 100568. https://doi.org/10.1016/j.pcorm.2025.100568

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. In *Psychological Methods* (Vol. 23, Number 3, pp. 412–433). American Psychological Association. https://doi.org/10.1037/met0000144

Milin, R., Kutcher, S., Lewis, S. P., Walker, S., Wei, Y., Ferrill, N., & Armstrong, M. A. (2016). Impact of a Mental Health Curriculum on Knowledge and Stigma Among High School Students: A Randomized Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(5), 383-391.e1. https://doi.org/10.1016/j.jaac.2016.02.018

Nobre, J., Oliveira, A. P., Monteiro, F., Sequeira, C., & Ferré-Grau, C. (2021). Promotion of Mental Health Literacy in Adolescents: A Scoping Review. In *International Journal of Environmental Research and Public Health* (Vol. 18, Number 18, p. 9500). https://doi.org/10.3390/ijerph18189500

Nuha, M. S., Indreswari, H., & Bariyyah, K. (2024). Psycho-educational group leaders through the homeroom approach: A systematic literature review. *ProGCouns: Journal of Professionals in Guidance and Counseling*, *5*(1 SE-Articles), 21–32. https://doi.org/10.21831/progcouns.v5i1.72261

Pieper, Martin, Roelle, Julian, vom Hofe, Rudolf, Salle, Alexander, & Berthold, Kirsten. (2021). Feedback in Reflective Journals Fosters Reflection Skills of Student Teachers. *Psychology Learning & Teaching*, *20*(1), 107–127. https://doi.org/10.1177/1475725720966190

Radez, J., Reardon, T., Creswell, C., Lawrence, P. J., Evdoka-Burton, G., & Waite, P. (2021). Why do children and adolescents (not) seek and access professional help for their mental health

problems? A systematic review of quantitative and qualitative studies. *European Child & Adolescent Psychiatry*, *30*(2), 183–211. https://doi.org/10.1007/s00787-019-01469-4

Renwick, L., Pedley, R., Johnson, I., Bell, V., Lovell, K., Bee, P., & Brooks, H. (2024). Mental health literacy in children and adolescents in low- and middle-income countries: a mixed studies systematic review and narrative synthesis. *European Child & Adolescent Psychiatry*, *33*(4), 961–985. https://doi.org/10.1007/s00787-022-01997-6

Sohal, M., Singh, P., Dhillon, B. S., & Gill, H. S. (2022). Efficacy of journaling in the management of mental illness: a systematic review and meta-analysis. *Family Medicine and Community Health*, *10*(1), e001154. https://doi.org/10.1136/fmch-2021-001154

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J. B., Jones, P., Kim, J. H., Kim, J. Y., Carvalho, A. F., Seeman, M. V, Correll, C. U., & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295. https://doi.org/10.1038/s41380-021-01161-7

Soria-Martínez, M., Navarro-Pérez, C. F., Pérez-Ardanaz, B., & Martí-García, C. (2024). Conceptual framework of mental health literacy: Results from a scoping review and a Delphi survey. *International Journal of Mental Health Nursing*, *33*(2), 281–296. https://doi.org/10.1111/inm.13249

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Uysal, G., Semerci, R., Şengün, R., & Düzkaya, D. S. (2025). Turkish validity and reliability of the universal mental health literacy scale for adolescents (10-14 years): A methodological study. *Journal of Evaluation in Clinical Practice*, *31*(5), e14171. https://doi.org/10.1111/jep.14171

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Contributors, S. 1. . (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wahdi, A. E., Astrini, Y. P., Setyawan, A., Fine, S. L., Ramaiya, A., Li, M., Wado, Y. D., Loi, V. M., Maravilla, J. C., Scott, J. G., Wilopo, S. A., & Erskine, H. E. (2025). Mental health service use among adolescents in three low- and middle-income countries: An analysis of the National Adolescent Mental Health Surveys. *Child and Adolescent Psychiatry and Mental Health*, *19*(1), 84. https://doi.org/10.1186/s13034-025-00924-2

Wang, Q., Wang, Q., Ji, Y., Chen, K., Li, K., Jia, F., & Peng, T. (2024). Construction and Validity of Chinese Translation of the Universal Mental Health Literacy Scale for Adolescents. In *International Journal of Mental Health Promotion* (Vol. 26, Number 8). https://doi.org/10.32604/ijmhp.2024.053127

Wang, Y., Li, M., Xu, H., He, F., Li, M., Wang, L., & Zang, Y. (2026). Trainee-delivered school-based intervention with guided narrative writing for adolescent anxiety and depression: a cluster randomized controlled trial. *BMC Medicine*, *24*(1), 289. https://doi.org/10.1186/s12916-026-04816-w

Xu, Z., Huang, F., Kösters, M., Staiger, T., Becker, T., Thornicroft, G., & Rüsch, N. (2018). Effectiveness of interventions to promote help-seeking for mental health problems: systematic review and meta-analysis. *Psychological Medicine*, *48*(16), 2658–2667. https://doi.org/DOI: 10.1017/S0033291718001265