



Konstruksi Instrumen Analisis Muatan Penalaran Matematis pada Buku Teks Matematika

Caesarla Elintang Yolawati¹, Raden Rosnawati^{2*}

¹ Mahasiswa Pendidikan Matematika, FMIPA, Universitas Negeri Yogyakarta, Yogyakarta

² Pendidikan Matematika, FMIPA, Universitas Negeri Yogyakarta, Yogyakarta

Caesarla0233fmipa.2022@student.uny.ac.id

Abstract

This study aims to construct an analysis instrument for deductive and inductive reasoning content in mathematics textbooks. This study uses a mix-method approach with the following stages: 1) preliminary study; 2) design and review of the test instrument; 3) testing the test instrument. The validity of the instrument was tested through expert judgment involving 2 mathematics education lecturers. Reliability was tested through intra-rater and inter-rater reliability. Intra-rater reliability was through repeated analysis by the researcher with an interval of 2 weeks, while inter-rater reliability involved 3 mathematics teachers. The construction of the reasoning analysis instrument for mathematics textbooks contains 2 aspects, namely inductive reasoning and deductive reasoning, each consisting of 4 indicators, which were declared valid and suitable for use after revision. The results of intra-rater reliability showed a high level of reliability and several indicators achieved perfect reliability and the results of inter-rater reliability have met a good level of agreement between the researcher and the rater. Thus, the developed instrument is considered valid and reliable for analyzing the content of deductive and inductive reasoning in mathematics textbooks.

Keywords: deductive reasoning; inductive reasoning; mathematics textbooks; validity; reliability.

Abstrak

Penelitian ini bertujuan untuk mengonstruksi instrumen analisis muatan penalaran deduktif dan induktif pada buku teks matematika. Penelitian ini menggunakan pendekatan mix-methods dengan tahapan: 1) studi pendahuluan; 2) perancangan dan penelaahan instrumen tes; 3) pengujian instrumen tes. Validitas instrumen diuji melalui expert judgement yang melibatkan 2 dosen pendidikan matematika. Reliabilitas diuji melalui reliabilitas intra-rater dan inter-rater. Reliabilitas intra-rater melalui analisis berulang peneliti dengan selang waktu 2 minggu, sedangkan reliabilitas inter-rater melibatkan 3 guru matematika. Konstruksi instrumen analisis penalaran buku teks matematika memuat 2 aspek yaitu penalaran induktif dan penalaran deduktif yang masing-masing terdiri dari 4 indikator dinyatakan valid dan layak digunakan setelah revisi. Hasil reliabilitas intra-rater menunjukkan tingkat reliabilitas tinggi dan beberapa indikator mencapai reliabilitas sempurna dan hasil reliabilitas inter-rater telah memenuhi tingkat kesepakatan yang baik antara peneliti dan rater. Dengan demikian, instrumen yang dikembangkan dinilai valid dan reliabel untuk menganalisis muatan penalaran deduktif dan induktif pada buku teks matematika.

Kata Kunci: penalaran deduktif, penalaran induktif, buku teks matematika, validitas, reliabilitas

1. PENDAHULUAN

Berdasarkan Risalah Kebijakan menurut Pusat Standar dan Kebijakan Pendidikan (2022) sebagian besar sekolah memberikan anggaran untuk pembelian buku teks Kurikulum Merdeka dan menjadi pilihan hampir seluruh sekolah. Berdasarkan hasil survey, sebanyak 93,22% responden kepala satuan pendidikan mengatakan buku teks cetak sebagai perangkat yang lebih dipilih dibandingkan dengan buku digital. Beberapa daerah di antaranya Jawa, Kalimantan, Maluku, Nusa Tenggara, Papua, Sulawesi, dan Sumatera, total penganggaran pada pembelian buku teks Kurikulum Merdeka dari mencapai 61,40%. Selain itu total pengetahuan mengenai kebutuhan buku teks Kurikulum Merdeka adalah 43,33%, pembelian buku teks yang memenuhi rasio kecukupan siswa dan guru adalah 58,19% dan 87,29%. Oleh karena itu, penggunaan buku teks di Indonesia cukup tinggi dan sebagian besar sekolah sudah menggunakan buku teks Kurikulum Merdeka.

Buku teks menerjemahkan kebijakan kurikulum suatu negara ke dalam representasi tersebut. Buku teks dimaksudkan sebagai mediator antara tujuan perancang kebijakan kurikulum dan guru yang memberikan pengajaran di kelas. Hal ini menunjukkan bahwa buku teks memiliki pengaruh yang kuat terhadap apa yang terjadi di dalam kelas (Valverde et al., 2002). Dockx et al (2020) mengatakan bahwa peran buku teks dalam cara siswa diajarkan biasanya dijelaskan dengan terlebih dahulu membedakan antara *intended curriculum* (kurikulum yang diharapkan), *implemented curriculum* (kurikulum yang diimplementasikan), dan *attained curriculum* (kurikulum yang dicapai). *Intended curriculum* adalah hal-hal yang seharusnya dipelajari oleh siswa berdasarkan rencana pemerintah. *Implemented curriculum* adalah hal-hal yang diajarkan guru setiap hari kepada siswanya di kelas. Siswa belajar dari *implemented curriculum*, tetapi kompetensi yang mereka capai, perilaku yang mereka lakukan, dan sikap yang mereka miliki mungkin berbeda dari yang dimaksudkan oleh guru itulah yang disebut sebagai *attained curriculum*. Akker (2010) mengatakan ranah *intended curriculum* mengacu pada pengaruh para pembuat kebijakan kurikulum dan pengembang kurikulum (dalam berbagai peran). *Implemented curriculum* khususnya berkaitan dengan dunia sekolah dan guru. *Attained curriculum* berkaitan dengan siswa.

Buku teks menerjemahkan tujuan resmi menjadi panduan bagi guru dengan materi yang dapat digunakan sehari-hari. Namun, guru tetap dapat membuat keputusan sendiri tentang cara menggunakan buku teks. Guru dapat memilih bagian-bagian buku teks dan menentukan berapa banyak waktu yang dihabiskan untuk topik-topik tertentu dalam buku teks. Oleh karena itu, buku teks berupa "*potentially implemented curriculum*" atau "kurikulum yang berpotensi diimplementasikan". Hal tersebut karena buku teks juga merupakan materi pendidikan utama yang dipelajari siswa di kelas, buku teks juga menjadi perantara antara *implemented curriculum* dan *attained curriculum*.

Utari dan Hartono (2019) mengatakan bahwa buku berfungsi sebagai fasilitator yang mampu membantu siswa belajar mandiri mengembangkan kemampuannya termasuk mengembangkan kemampuan penalaran siswa. Buku teks hendaknya memuat bahan bacaan yang memberikan kesempatan bagi siswa untuk mengembangkan kemampuan

penalaran dan pembuktian matematis. Oleh karena itu, di dalam buku teks tersebut perlu memuat materi maupun soal latihan yang mengandung muatan penalaran matematis. Hal tersebut bertujuan untuk melatih kemampuan penalaran matematis siswa, khususnya penalaran deduktif dan induktif.

Standar *National Council for Teachers of Mathematics* (NCTM) (2004) menyebutkan bahwasannya salah satu standar proses adalah penalaran dan pembuktian (*reasoning and proof*). Dua aspek penalaran matematis yang mendefinisikan soal PISA dan sangat penting bagi dunia saat ini adalah penalaran deduktif yang merupakan ciri khas dari proses matematika dan penalaran induktif. Kegunaan keduanya sangat ditekankan (OECD, 2018). Penalaran deduktif juga merupakan suatu proses ketika informasi baru diperoleh melalui seperangkat premis yang dapat dibuktikan dengan alasan rasional dan menghasilkan sebuah kesimpulan yang valid (Harel & Weber, 2020). Deduktif menyimpulkan klaim dari data dan memainkan penting atas proses pembuktian formal (Jeannotte & Kieran, 2017). Argumen deduktif yang valid menjaga kebenaran, dalam arti jika premis-premisnya benar, maka kesimpulannya juga benar (Ayalon dan Even, 2010).

Penalaran induktif mengidentifikasikan suatu pola atau variasi kasus yang merupakan tahap awal untuk memperoleh kesimpulan umum. Proses penalaran induktif pada penelitian ini terbagi menjadi tiga, yaitu: pengamatan terhadap suatu keteraturan, pembentukan suatu pola, perumusan generalisasi (Sosa-Moguel et al., 2019). Oleh karena itu berdasarkan pemaparan di atas, penalaran deduktif adalah proses penarikan kesimpulan yang valid melalui pembuktian dari aturan yang sudah benar sedangkan penalaran induktif adalah memiliki keteraturan, memiliki pola untuk diidentifikasi dan diverifikasi, kesimpulannya bersifat umum. Jadi dapat disimpulkan dari karakteristik penalaran deduktif yakni membuktikan suatu data dengan aturan matematika formal untuk mencapai kesimpulan yang valid, sedangkan penalaran induktif mengarah kepada dugaan terhadap pola-pola agar mencapai suatu keteraturan dan menghasilkan kesimpulan yang umum.

Instrumen yang baik perlu melalui uji validitas dan reliabilitas. Instrumen yang tepat dibutuhkan untuk proses pengukuran yang tepat, sehingga instrumen yang tepat dapat meminimalkan kesalahan pengukuran (Ulfah, et al., 2020). Jika penelitian harus sah, data yang mendasarinya, individu yang dilibatkan, dan proses yang dihasilkan, keseluruhannya harus handal. Keandalan (reliabilitas) menilai sejauh mana suatu desain penelitian dan data yang dihasilkan mewakili kondisi yang sebenarnya diteliti, kekhasan para analis secara individu dan bias tersembunyi dari sebuah prosedur. Reliabilitas merupakan sesuatu yang dibutuhkan, meskipun keandalan saja belum cukup untuk mengukur suatu instrumen, selain handal (reliable) instrumen juga harus sah (valid). Kesahihan (validitas) menggambarkan kualitas hasil penelitian yang diakui sebagai fakta yang tidak dapat ditentang (Krippendorff, 1993).

Beberapa penelitian terdahulu telah menganalisis penalaran matematis pada buku teks matematika menggunakan instrumen analisis yang masih bersifat umum, sehingga belum secara khusus mengkaji penalaran deduktif dan induktif. Oleh karena itu masih perlu dikembangkan instrumen untuk menganalisis muatan penalaran matematis yang

secara khusus berfokus pada penalaran deduktif dan induktif. Pengembangan instrumen analisis penalaran deduktif dan induktif karena kedua jenis penalaran tersebut memiliki karakteristik yang berbeda sehingga memerlukan indikator yang spesifik berfokus berdasarkan karakteristik masing-masing penalaran sehingga dapat membantu proses analisis penalaran secara lebih spesifik, sistematis, dan eksplisit. Instrumen ini dapat membantu memperjelas dan menunjukkan secara eksplisit letak serta karakteristik penalaran deduktif dan induktif yang terdapat pada materi maupun latihan soal. Dengan demikian, hasil analisis yang diperoleh diharapkan dapat membantu dalam pengembangan kualitas buku ajar matematika untuk memfasilitasi siswa agar melatih kemampuan penalarannya secara optimal.

2. METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan *mix methode* dengan tahapan: 1) studi pendahuluan meliputi survey lapangan serta studi literatur; 2) perancangan dan penelaahan instrumen tes yang memuat analisis dimensi muatan penalaran dalam buku teks, pemilihan indikator untuk masing-masing dimensi; 3) pengujian instrumen tes, yang melibatkan 13 buku teks matematika yang dianalisis terdiri dari buku teks matematika Indonesia kelas 7, 8, 10, 11, dan 11 lanjutan, serta buku teks sekolah menengah Singapura 1A, 1B, 2A, 2B, 3A, 3B, 4A, dan 4B. Validasi instrumen pada penelitian ini adalah validasi konstruksi yang dapat dilakukan melalui pendapat para ahli (*expert judgement*) yaitu dosen pendidikan matematika. Reliabilitas pada instrumen menggunakan reliabilitas intra-rater dan reliabilitas inter-rater. Reliabilitas intra-rater adalah apabila data diukur sejauh mana pun menggunakan prosedur pengukuran yang sama akan mendapatkan hasil yang sama pada percobaan berulang yang dilakukan oleh peneliti sendiri. Reliabilitas inter-rater adalah proses analisis yang dilakukan sejauh mana pun pada kondisi yang berbeda, lokasi yang berbeda, dan peneliti lain akan mendapatkan hasil yang sama (Krippendorff, 2004). Pada tahapan ini analisis dilakukan oleh peneliti dan 3 guru matematika.

Keputusan reliabilitas diekspresikan secara numerik dalam bentuk koefisien yang besarnya $-1,00 \leq \rho \leq 1,00$, semakin tinggi koefisien menunjukkan semakin tinggi reliabilitasnya, indeks koefisien reliabilitasnya minimal sebesar 0,85 (Retnawati, 2015). Hal ini diterapkan pada reliabilitas Intra-rater. Pada reliabilitas interrater menggunakan perhitungan berdasarkan Krippendorff Alpha dengan interpretasinya (Marzi et al, 2024). Interpretasi Krippendorff Alpha dapat dilihat dalam Tabel 1.

Tabel 1. Interpretasi Krippendorff Alpha

Alpha	Interpretasi
1	Sempurna
$\geq 0,80$	Memuaskan
$0,67 \leq \alpha \leq 0,79$	Cukup
$< 0,67$	Kurang
0	Tidak ada kesepakatan
< 0	Perbedaan Pendapat

3. HASIL DAN PEMBAHASAN

Pada penelitian ini kerangka kerja analisis muatan penalaran matematis mengkhususkan penalaran deduktif dan induktif. Kerangka kerja ini didasarkan pada pengembangan dari kerangka kerja Otten et al (2014), menganalisis 6 buku teks geometri tingkat menengah Amerika Serikat untuk mencirikan pembenaran yang dilakukan pada kegiatan eksposisi dan kegiatan penalaran dan pembuktian yang diharapkan siswa melalui latihan soal. Melalui kerangka kerja penalaran dan pembuktian matematika berfokus pada geometri tingkat menengah dan analisis mencakup bagian eksposisi (materi) dan latihan siswa. Selanjutnya penelitian Thompson et al (2012) membahas sifat dan tingkat penalaran dan pembuktian dalam kurikulum pada 20 buku teks matematika sekolah menengah kontemporer Amerika Serikat. Melalui analisis terhadap 20 buku teks dari enam seri berbeda di Amerika Serikat, penelitian ini menemukan variasi dalam penekanan pada pengembangan argumen, evaluasi argumen, dan pembuktian.

Penelitian ini juga menunjukkan bahwa ada kebutuhan untuk lebih banyak penelitian yang mengeksplorasi bagaimana penalaran terkait bukti dapat lebih efektif diintegrasikan ke dalam materi pembelajaran matematika untuk meningkatkan kemampuan siswa dalam mengembangkan dan mengevaluasi argumen matematika. Penelitian ini menggunakan kerangka kerja yang didasarkan pada Standar Penalaran dan Pembuktian NCTM dan kerangka kerja TIMSS untuk menganalisis buku teks, mengembangkan kerangka kerja analisis sendiri untuk membedakan antara dugaan, argumen, dan bukti dalam pembelajaran matematika. Penelitian Zhang dan Qi (2019) meneliti muatan penalaran dan pembuktian pada buku teks matematika kelas 8 yang berorientasi reformasi di Tiongkok serta membandingkannya dengan buku teks Amerika Serikat. Studi ini mengeksplorasi buku teks yang menyajikan peluang siswa untuk belajar penalaran dan pembuktian dengan membandingkan penyebaran dan jenis penalaran dan pembuktian dalam berbagai volume buku serta konten. Kerangka kerja pada penelitian ini mengkategorikan dan menganalisis tugas-tugas dan narasi dalam buku teks berdasarkan jenis penalaran dan pembuktian yang disajikan, termasuk membuat dugaan, memberikan bukti, dan evaluasi argumen.

Penelitian Utari dan Hartono (2019) mendeskripsikan muatan penalaran dan pembuktian matematis disajikan dalam buku teks matematika untuk siswa kelas X Kurikulum 2013 di Indonesia. Hasil penelitiannya menunjukkan bahwa terdapat peluang untuk siswa dalam mengembangkan kemampuan penalaran dan pembuktian melalui bacaan justifikasi, tetapi apabila diperiksa melalui evaluasi ditemukan bahwa belum ada peluang yang diberikan untuk siswa untuk melakukan penalaran dan pembuktian melalui penyelesaian soal yang berkaitan dengan indikator untuk mengevaluasi argumen siswa secara mendalam.

3.1 Instrumen Analisis Muatan Penalaran Matematis pada Buku Teks Matematika

Kerangka kerja analisis muatan penalaran matematis yang dikembangkan didasarkan pada penalaran matematis membutuhkan pemahaman implisit kesimpulan

induktif dan argumen deduktif (Morris, 2002). Selanjutnya penalaran matematis dibedakan menjadi dua yaitu penalaran induktif dan penalaran deduktif, oleh sebab itu fokus analisis muatan penalaran matematika yang dikembangkan adalah penalaran induktif dan deduktif. Kajian muatan penalaran matematis pada buku memfokuskan pada menganalisis penyajian materi dan latihan soal. Penalaran deduktif terdiri dari 4 indikator yaitu pernyataan, pernyataan penalaran deduktif, menyelidiki kebenaran, membuktikan secara formal. Pada indikator pernyataan, kajian muatan penalaran deduktif untuk buku pada penyajian materi terbagi menjadi dua yaitu pernyataan umum dan pernyataan khusus, sedangkan untuk menganalisis latihan soal terbagi menjadi 3 yaitu pernyataan umum, pernyataan khusus, dan pernyataan umum dengan contoh khusus. Penalaran induktif terdiri dari 4 indikator yang jumlahnya sama untuk menganalisis penyajian materi dan latihan soal di antaranya mengamati pola, memprediksi pola, generalisasi, dan justifikasi. Secara rinci kerangka kerja instrumen disajikan pada Tabel 2.

Tabel 2. Kerangka Kerja Instrumen Analisis Muatan Penalaran Buku Teks

Aspek	Struktur Buku	Definisi Operasional
Penalaran deduktif dalam buku teks	Materi	Penalaran deduktif dalam materi sajian yang disajikan melalui tahapan dengan aturan dan premis-premis yang dipercayai kebenarannya untuk mendapatkan kesimpulan yang valid
	Latihan soal	Penalaran deduktif dalam latihan soal disajikan berupa soal yang memuat stimulus berupa pernyataan menuntut penyelesaian melalui tahapan menggunakan aturan dan premis-premis yang dipercayai kebenarannya untuk mendapatkan kesimpulan yang valid
Penalaran induktif dalam buku teks	Materi	Penalaran induktif dalam materi merupakan sajian materi yang mengarahkan pada pola dan keteraturan untuk mencapai kesimpulan umum yang dapat diuji kebenarannya melalui percobaan
	Latihan soal	Penalaran induktif dalam latihan soal disajikan berupa soal yang memuat stimulus berupa pernyataan yang mengarahkan kepada pola dan keteraturan, menuntut penyelesaian melalui pola dan keteraturan untuk mencapai kesimpulan umum yang dapat diuji kebenarannya melalui percobaan

Berdasarkan kerangka kerja di atas dikembangkan instrumen analisis muatan penalaran matematis, dimana penalaran deduktif terdiri dari 4 indikator yaitu pernyataan, pernyataan penalaran deduktif, menyelidiki kebenaran, membuktikan secara formal, sedangkan untuk penalaran induktif terdiri dari 4 indikator yaitu mengamati pola, memprediksi pola, generalisasi dan justifikasi. Secara rinci indikator disajikan pada Tabel 3 berikut.

Tabel 3. Kisi-Kisi Instrumen Analisis Muatan Penalaran Buku Teks

Aspek	Indikator	Deskriptor
Penalaran deduktif dalam buku teks	Pernyataan yang bernilai benar	Klaim atau kalimat dalam matematika yang menyatakan suatu kondisi, hubungan, atau sifat dari objek-objek matematis. Pernyataan ini bisa berupa hipotesis, teorema, definisi, atau kesimpulan, yang memiliki kebenaran yang dapat diuji atau dibuktikan. Pernyataan terbagi menjadi 2, yaitu:

Aspek	Indikator	Deskriptor
Pernyataan penalaran deduktif		1. Pernyataan umum adalah pernyataan yang menyangkut seluruh kelas objek atau situasi matematika.
		2. Pernyataan khusus adalah pernyataan yang menyangkut objek atau situasi matematika
		3. Pernyataan umum dengan contoh khusus adalah pernyataan yang menyangkut seluruh kelas objek matematika tetapi terdapat anggota kelas tertentu yang telah ditunjukkan.
		Memuat kode khusus yang mengarah kepada penalaran deduktif berupa kode yang bersifat eksplisit dan kode yang bersifat implisit. 1. Kode bersifat eksplisit langsung menuliskannya secara gamblang, seperti contohnya “argumen deduktif”, “buktikan secara deduktif”, dsb. 2. Kode bersifat implisit tidak menuliskan secara terang-terangan seperti “buktikan”, “simpulkan”, “tunjukkan”, dan “jelaskan mengapa”(jika diikuti dengan menyelidiki kebenaran)
Penalaran induktif dalam buku teks	Menyelediki kebenaran	Menentukan nilai kebenaran dari klaim matematika untuk menilai klaim matematika tersebut benar atau salah dengan pembuktian formal, atau terdapat kalimat “benar atau salah”.
	Membuktikan secara formal	Menyertakan definisi, teorema, postulat, dan klaim lainnya yang didukung oleh bukti atau argumen logis yang telah disepakati dan diakui kebenarannya.
	Mengamati pola	Berupa pernyataan yang memuat gambar, angka, benda, atau objek yang memiliki atribut sama dan secara eksplisit atau implisit mengandung pola
	Memprediksi pola	Berupa kemungkinan-kemungkinan pola dan keteraturan yang terbentuk.
	Generalisasi	Berupa rumus umum atau bentuk umum dari keteraturan pola
	Justifikasi	Memvalidasi rumus umum dengan menggunakan kasus-kasus tertentu.

3.2 Validitas Instrumen

Pengujian validasi konstruksi dapat dilakukan melalui pendapat para ahli (expert judgement). Para ahli melakukan penilaian di setiap item untuk menentukan item tersebut mewakili domain yang diukur. Penilaian dilakukan oleh ahli yang tidak terlibat dalam pengembangan instrumen secara independen untuk mengurangi potensi bias. Proses penilaian dilakukan secara sistematis untuk mendapatkan hasil evaluasi yang objektif dan dapat dipercaya (Boateng et al., 2018). Pada penelitian ini, validitas dilakukan oleh ahli yaitu dosen pendidikan matematika yang memiliki kompetensi dalam bidang pembelajaran matematika dan analisis buku teks sebagai validator sebanyak 2 ahli. Pengujian validasi ini dilakukan untuk mengetahui kelayakan instrumen berdasarkan 3 aspek, yaitu petunjuk penggunaan instrumen, isi, dan bahasa. Tahap validasi terdiri dari 3 tahapan, yaitu: 1) peneliti menyusun instrumen muatan

penalaran matematis, 2) instrumen muatan penalaran matematis dikonsultasikan dan dievaluasi oleh validator untuk mendapatkan masukan dan perbaikan; 3) peneliti memperbaiki instrumen. Data validasi instrumen diperoleh dari hasil validasi ahli.

Berdasarkan hasil validasi dari 2 ahli sebagai validator, instrumen analisis penalaran deduktif dan induktif dinyatakan valid dan layak digunakan dengan revisi. Selain instrumen yang dinyatakan valid melalui tanda centang yang diberikan pada lembar validasi, terdapat pula data kualitatif yang diperoleh dari validator ahli berupa saran dan masukan yang digunakan sebagai perbaikan dalam pengembangan instrumen analisis muatan penalaran deduktif dan induktif pada buku teks matematika. Saran yang diberikan oleh kedua validator tersebut dapat dilihat pada Tabel 4.

Tabel 4. Komentar dan Saran Validator Terhadap Instrumen

Validator	Komentar dan Saran
Validator 1	1. Deskriptor dari beberapa indikator pada penalaran deduktif memuat beberapa kata yang sama sehingga perbedaan antar indikator belum jelas. 2. Dugaan di penalaran deduktif dibedakan dengan dugaan di penalaran induktif atau langsung dituliskan saja yang membedakannya agar lebih jelas. 3. Poin kategori “interaksi siswa dengan materi pembuktian” dipertimbangkan kembali karena setelah dilihat dari contoh, lebih mengarah kepada latihan soal dan sebaiknya dimasukkan ke dalam struktur buku untuk latihan soal saja. 4. Poin kategori kesimpulan hasil pada indikator penalaran deduktif sebaiknya dihapus saja karena masih tergolong proses dari menyelidiki dugaan dan tidak ada yang menunjukkan secara khusus bahwa satu soal tertentu masuk ke dalam kesimpulan umum saja. 5. Untuk kategori pernyataan pada penalaran deduktif pada penyajian materi dan latihan soal ditambahkan ke indikator 6. Poin lain seperti membenaran empiris, tidak ada membenaran, dan contoh tandingan sebaiknya dihapus saja karena fokus analisis untuk menganalisis penalaran deduktif, sedangkan ketiga poin tersebut tidak untuk menganalisis penalaran deduktif. 7. Poin mengamati kasus atau permasalahan pada penalaran induktif dihapuskan saja karena dilihat dari deskriptor sama dengan memprediksi pola
Validator 2	1. Perbaiki salah pengetikan dan kalimat yang kurang jelas atau kurang lengkap. 2. Perbaiki pernyataan yang ambigu pada instrumen penilaian muatan penalaran matematis.

3.3 Reliabilitas Instrumen

Reliabilitas menunjukkan sejauh mana penelitian dapat diulang dengan hasil yang konsisten (Cole, 2024). Reliabilitas yang dilakukan adalah reliabilitas intra-rater dan inter-rater. Reliabilitas intra-rater dilakukan dengan cara peneliti membaca, mengamati, mengategorikan muatan pada buku, dan menganalisis buku secara berulang untuk mendapatkan hasil yang sama. Jarak waktu antara analisis pertama dan kedua sejauh 2 minggu. Reliabilitas inter-rater dilakukan melalui kerja sama antara peneliti dengan 3 guru matematika untuk bersama-sama menganalisis buku. Cole (2024) mengatakan bahwa konsistensi antar rater secara kuantitatif, proses penyelarasan perbedaan dari

hasil analisis untuk mencapai kesepakatan dalam menyimpulkan data melalui kesepakatan bersama. Peneliti yang dilibatkan 2 atau lebih secara independen menganalisis data yang sama dan kemudian membandingkan hasilnya untuk meminimalkan perspektif selektif dan bias interpretasi. Sebelum guru menganalisis, terlebih dulu guru diberikan pemahaman terkait penalaran deduktif dan induktif serta langkah dalam menganalisis buku teks melalui indikator penalaran deduktif dan induktif, contoh penalaran deduktif dan induktif di buku teks matematika, dan lembar analisis yang disusun. Berikutnya guru menganalisis buku teks SMP dan SMA secara independen.

Hasil reliabilitas intra-rater pada buku teks Indonesia dan Singapura, sedangkan hasil reliabilitas inter-rater untuk masing-masing 1 bab pada penalaran deduktif dan induktif buku teks Indonesia. Hasil reliabilitas untuk penalaran deduktif dijabarkan dalam Tabel 5.

Tabel 5. Hasil Reliabilitas Penalaran Deduktif Buku Teks Indonesia dan Singapura

Reliabilitas		Indonesia				Singapura		
		Intra-rater		Inter-rater		Intra-rater		
		Materi	Latihan Soal	Materi	Latihan Soal	Materi	Latihan Soal	
Indikator Muatan Deduktif	Pernyataan	Umum	0,883	0,933	1	1	-	1
		Khusus	0,779	0,989	0,726	0,8	1	0,989
		Umum Contoh Khusus	1	-	0,24	0,663	1	1
	Pernyataan Penalaran Deduktif	Eksplisit	-	-	1	1	-	-
		Implisit	0,957	0,992	0,736	0,78	1	0,99
	Menyelidiki Kebenaran		0,957	0,992	0,683	0,774	1	0,99
	Membuktikan Formal		0,942	0,992	0,688	0,811	1	0,99

Pada reliabilitas intra-rater terdapat beberapa indikator yang tidak memiliki nilai reliabilitas, hal tersebut dikarenakan tidak ditemukan muatan penalaran deduktif pada indikator tersebut saat analisis pertama dan analisis kedua. Hasil reliabilitas penalaran induktif dijabarkan dalam Tabel 6.

Tabel 6. Hasil Reliabilitas Penalaran Induktif pada Buku Teks Indonesia dan Singapura

Negara		Indonesia				Singapura	
Reliabilitas		Intra-rater		Interrater		Koefisien Internal	
		Materi	Latihan Soal	Materi	Latihan Soal	Materi	Latihan Soal
Indikator Penalaran Induktif	Mengamati Pola	1	0,987	0,442	0,338	0,994	0,995
	Memprediksi Pola	0,99	0,987	0,402	0,386	0,994	0,995
	Generalisasi	1	0,997	0,828	0,791	0,971	0,98
	Justifikasi	1	0,752	0,794	0,213	0,985	0,996

Uji reliabilitas pertama yang dilakukan untuk mengetahui konsistensi hasil pengukuran instrumen. Pada tahapan ini, peneliti membaca, mengamati, mengkategorikan muatan penalaran deduktif dan induktif pada buku, serta menganalisis buku teks secara berulang untuk mendapatkan hasil yang sama. Jarak waktu analisis pertama dan kedua dilakukan dalam selang waktu 2 minggu. Pelaksanaan kedua analisis dengan memberikan jeda waktu diberikan agar hasil analisis kedua tidak terpengaruh oleh ingatan peneliti terhadap hasil analisis pertama sehingga dapat menunjukkan kestabilan instrumen secara lebih objektif. Penemuan tinjauan literatur pada penelitian menunjukkan bahwa rekomendasi jangka waktu yang diberikan untuk reliabilitas tidak selalu konsisten. Selain itu panduan umum yang ditawarkan adalah jangka waktu harus cukup lama untuk menghindari efek ingatan sebelumnya tetapi cukup singkat juga agar perubahan yang disebabkan oleh peningkatan kemampuan di bidang tersebut dapat diminimalkan (Wyse, 2021). Penentuan interval waktu perlu dipertimbangkan dalam desain penelitian reliabilitas. Hingga saat ini belum terdapat ketentuan lamanya waktu interval yang diberikan. Sebagai gambaran, interval waktu digunakan selama 2 minggu, tetapi tidak ada aturan standar yang berlaku (De Vet, 2011).

Penelitian-penelitian terdahulu telah mengkaji pengaruh interval waktu antara pengujian terhadap estimasi reliabilitas, secara empiris maupun meta-analisis. Hasil penelitian secara umum menunjukkan perubahan koefisien reliabilitas seiring dengan perbedaan interval waktu pada pengujian pertama dan kedua. Reliabilitas cenderung menurun semakin panjang interval waktu yang diberikan, meskipun pola tersebut tidak selalu konsisten (Wyse, 2021). Pada penelitian Wyse (2021) menunjukkan bahwa koefisien reliabilitas tertinggi diperoleh ketika tes kedua diberikan setidaknya 2 hingga 3 minggu setelah tes pertama. Pada tes membaca dan matematika yang diteliti, koefisien reliabilitas bentuk tertinggi diperoleh ketika peserta diuji ulang setelah menunggu 2 hingga 3 minggu. Koefisien seringkali sama setelah melewati jangka waktu tersebut. Penelitian Marx et al (2003) menyimpulkan bahwa tidak ada perbedaan signifikan secara

klinis ataupun statistik terhadap reliabilitas yang dilakukan antara 2 hari dibandingkan dengan interval 2 minggu. Berdasarkan pemaparan tersebut, penggunaan waktu interval selama 2 minggu dalam penelitian ini memadai karena penggunaan interval tersebut masih berada dalam waktu yang umum digunakan pada pengujian reliabilitas. Interval tersebut memberikan jeda yang cukup sehingga peneliti tidak mengingat secara rinci hasil analisis sebelumnya, tetapi pemahaman terhadap instrumen analisis masih terjaga secara konsisten. Dengan demikian analisis kedua yang dilakukan berdasarkan interpretasi isi buku sesuai indikator instrumen, bukan berdasarkan ingatan pada analisis sebelumnya. Selain itu, buku teks matematika yang merupakan objek pada penelitian ini bersifat tetap serta tidak mengalami perubahan selama proses penelitian. Oleh karena itu interval 2 minggu yang digunakan diharapkan meminimalkan ingatan pada analisis pertama tanpa dipengaruhi oleh perubahan pada objek yang dianalisis.

Menurut Retnawati (2015) Keputusan reliabilitas disajikan secara numerik dalam bentuk koefisien yang besarnya $-1,00 \leq \rho \leq 1,00$, semakin tinggi koefisien menunjukkan semakin tinggi reliabilitasnya, dengan indeks koefisien reliabilitasnya minimal sebesar 0,85. Berdasarkan hasil reliabilitas yang dilakukan intra-rater, untuk semua indikator penyajian materi penalaran deduktif mendapatkan hasil $>0,85$ kecuali indikator pernyataan untuk sub indikator pernyataan khusus mendapatkan hasil $<0,85$ tepatnya 0,779. Namun meskipun mendapatkan nilai di bawah 0,85 hasil intra-rater tersebut masih tergolong reliable, karena berdasarkan penelitian Nguyen et al (2018) mengatakan bahwa pengukuran reliabilitas yang biasanya dinilai dengan koefisien yang berkisar antara -1 (korelasi negatif sempurna) hingga 1 (korelasi positif sempurna) nilai reliabilitasnya tergolong tinggi lebih disukai jika nilai korelasinya $\geq 0,70$. Selain itu nilai konsistensi reliabilitas pada semua indikator penalaran deduktif untuk latihan soal pada buku teks indonesia $>0,85$ dan pada materi serta latihan soal di buku teks Singapura juga ada beberapa indikator yang mendapatkan nilai 1. Oleh karena itu artinya hasil reliabilitas Intra-rater tergolong sempurna. Hasil reliabilitas intra-rater penalaran induktif pada buku teks indonesia untuk penyajian materi dan latihan soal mendapatkan hasil $>0,85$ dan indikator mengamati pola, generalisasi, dan justifikasi mendapatkan hasil koefisien 1 yang berarti reliabilitasnya sempurna. Namun, pada indikator justifikasi untuk latihan soal mendapatkan nilai $<0,85$ tepatnya 0,752, tetapi berdasarkan penelitian yang dilakukan oleh Nguyen et al (2018) hasil koefisien tersebut mengartikan bahwa reliabilitas masih tergolong tinggi. Hasil reliabilitas penalaran induktif pada buku teks Singapura untuk materi dan latihan soal di semua indikator tergolong tinggi karena koefisiennya $> 0,85$.

Pada reliabilitas inter-rater terdapat beberapa indikator pada penalaran deduktif yang nilai reliabilitasnya tergolong kurang menurut perhitungan reliabilitas inter-rater Krippendorff Alpha karena nilai alpha berada di $0,67 \leq \alpha \leq 0,79$, di antaranya indikator pernyataan khusus pada penyajian materi, pernyataan umum dengan contoh khusus pada penyajian materi dan latihan soal, indikator pernyataan deduktif implisit pada penyajian materi dan latihan soal, indikator menyelidiki kebenaran pada pernyataan materi dan latihan soal, serta indikator membuktikan formal pada penyajian materi. Pada penalaran induktif juga ada indikator yang reliabilitasnya tergolong kurang, di

antaranya indikator penyajian materi pada mengamati pola pada penyajian materi dan latihan soal, memprediksi pola pada penyajian materi dan latihan soal, serta justifikasi pada latihan soal.

Meskipun sebelum melakukan analisis buku teks, sebelumnya para rater telah diberikan panduan analisis melalui instrumen yaitu indikator penalaran deduktif dan induktif serta contoh penalaran deduktif dan induktif di buku teks. Peneliti juga memberikan penjelasan kepada rater mengenai panduan analisis di instrumen tersebut untuk menyamakan pemahaman antar rater serta berdiskusi terkait poin-poin yang tidak dipahami oleh rater. Namun berdasarkan hasil reliabilitas masih terdapat indikator yang belum menunjukkan tingkat konsistensi yang rendah. Hal ini disebabkan karena terdapat perbedaan sudut pandang di antara para guru dengan peneliti dalam menginterpretasikan bentuk penalaran yang ditampilkan dalam buku teks. Perbedaan ini menunjukkan belum ada keseragaman persepsi secara optimal terhadap penalaran deduktif dan induktif yang disajikan dalam buku teks.

Variasi reliabilitas di beberapa indikator dapat dipengaruhi oleh beberapa faktor meskipun para rater telah diberikan pedoman analisis yang sama, tetapi terdapat kemungkinan dalam memahami dan menginterpretasikan pedoman yang diberikan. Perbedaan hasil analisis antar-rater tidak selalu menunjukkan kesalahan dalam proses penilaian melainkan menunjukkan variasi dan sudut pandang rater yang berbeda ketika menginterpretasikan objek yang dianalisis berdasarkan proses kognitif dan pengalaman para rater.

Pendukung untuk menjamin validitas penilaian adalah proses kognitif antar rater yang selaras. Namun bahwa perilaku kognitif antar rater bersifat fleksibel dan mudah dibentuk meskipun tetap mengacu pada pedoman penilaian yang digunakan. Secara umum, proses kognitif rater dianggap sebagai aspek yang penting selama proses penilaian. Penelitian ini menemukan bahwa terdapat 2 aspek utama pada proses kognitif rater yaitu memahami dan membentuk penilaian terhadap objek yang dinilai serta menentukan cara pemberian skor berdasarkan hasil penilaian tersebut. Setiap penilai memiliki pengalaman dan tingkat keahlian yang berbeda, sehingga menyebabkan perbedaan terhadap penilaian untuk menginterpretasikan objek yang dinilai. Oleh karena itu meskipun menggunakan komposisi penilaian yang sama, interpretasi yang dibangun oleh rater bisa saja berbeda karena para penilai memiliki cara masing-masing untuk menentukan skor terhadap objek yang sedang diuji, sehingga hasil penilaian yang diberikan dapat berbeda (Jia & Zhang, 2023).

Penelitian Lamprianou et al (2023) menunjukkan bahwa secara kualitatif pengalaman yang berbeda mempengaruhi keputusan penilaian. Pengalaman rater juga merupakan faktor penting yang dapat berpengaruh pada konsistensi, ketelitian, dan kualitas hasil penilaian. Namun masih terdapat variasi mengenai cara mengukur pengalaman penilai di beberapa penelitian sebelumnya, di antaranya pengalaman mengajar, pengalaman melakukan penilaian, keterlibatan proyek penilaian, pelatihan, dan gelar master. Cole (2024) mengatakan perbedaan yang terjadi di antara peneliti bisa saja terjadi karena pengalaman yang berbeda, pemahaman topik yang berbeda, dan interpretasi yang berbeda. Oleh karena itu proses rekonsiliasi atau proses diskusi

penyelarasan sebagai diskusi antar rater dan peneliti dapat dilakukan untuk membahas perbedaan pandangan.

Hasil uji reliabilitas inter-rater menunjukkan terdapat variasi kesepakatan di beberapa indikator penalaran deduktif dan induktif. Meskipun sebelum analisis dilakukan telah diberikan pengarahan oleh peneliti kepada rater, instrumen juga berisi deskriptor dan contoh pada indikator untuk menyamakan persepsi. Namun nilai reliabilitas yang muncul tergolong kurang dan menginterpretasikan perbedaan sudut pandang antara rater dan peneliti. Menurut Cole (2024) bahwa makna yang berbeda-beda merupakan bagian dari suatu proses yang memberikan informasi lebih besar untuk didiskusikan dan tidak mengurangi nilai dari data yang dianalisis. Diskusi dilakukan untuk menentukan kesepakatan yang sama. Melalui perhitungan statistik dapat membantu mengidentifikasi bagian yang menunjukkan perbedaan, sehingga diskusi dilakukan secara lebih fokus pada bagian yang berbeda.

Proses penyelarasan dilakukan dengan melibatkan seluruh anggota dan terdapat kemungkinan yang dilakukan : (1) membuat kode/indikator baru, (2) memasukkan suatu kode/indikator sebagai sub kategori dari indikator yang lebih umum, (3) menggabungkan beberapa kode/indikator menjadi satu yang lebih sesuai, atau (4) menghapus kode/indikator yang tidak diperlukan. Peneliti dapat memutuskan kode/indikator yang perlu disusun ulang. Suatu kode/indikator wajar jika mengalami perubahan, nama yang berganti, indikator baru yang dibuat, dihapus, digabungkan, atau dimasukkan menjadi sub bagian. Ada banyak cara untuk menunjukkan perubahan kode/indikator pada reliabilitas inter-rater. Pada penelitian ini setelah diperoleh hasil beberapa indikator yang tidak reliabel yang mengindikasikan terdapat perbedaan interpretasi antara peneliti dan rater meski pada indikator sudah diberikan penjelasan melalui deskriptor. Perbedaan ini bukan merupakan kesalahan melainkan peluang untuk mengidentifikasi indikator tersebut. Berdasarkan hasil tersebut, peneliti memperjelas deskriptor pada indikator-indikator yang tidak reliabel tersebut dan langkah yang diambil adalah pendefinisian ulang untuk mereduksi deskriptor yang menyebabkan interpretasi berbeda untuk menyelaraskan persepsi antar-rater lainnya di masa depan.

Berdasarkan pemaparan di atas maka secara keseluruhan instrumen analisis muatan penalaran matematis secara keseluruhan dinyatakan reliabel karena didukung oleh reliabilitas intra-rater yang menunjukkan tingkat sangat tinggi hingga sempurna. Pada reliabilitas inter-rater, perbedaan interpretasi tidak dipandang sebagai kesalahan melainkan bagian dari proses penyempurnaan instrumen untuk memperjelas deskriptor melalui pendefinisian ulang agar memberikan pemahaman yang seragam di masa depan terhadap penalaran deduktif dan induktif di buku teks. Sebagai contoh, deskriptor sebelumnya berpotensi menimbulkan perbedaan pendapat para rater, sehingga pada indikator yang menyebabkan perbedaan interpretasi tersebut diperjelas dengan memperjelas deskripsi indikator serta melengkapi contoh-contoh yang merepresentasikan indikator tersebut agar rater memiliki acuan yang lebih seragam untuk menganalisis buku teks matematika. Dengan demikian instrumen akhir yang dihasilkan tetap dapat digunakan secara reliabel sebagai pedoman analisis muatan penalaran matematis di buku teks.

4. SIMPULAN

Instrumen analisis muatan penalaran deduktif dan induktif yang terdiri dari aspek penalaran deduktif dan induktif pada buku teks matematika, telah memenuhi aspek validitas isi berdasarkan penilaian ahli dan dinyatakan layak digunakan setelah dilakukan revisi. Hasil uji reliabilitas terdiri dari reliabilitas intra-rater dan reliabilitas inter-rater. Hasil uji reliabilitas intra-rater menunjukkan tingkat konsistensi yang tinggi hingga sempurna, dan uji reliabilitas inter-rater menunjukkan bahwa secara umum sudah terdapat kesepakatan terhadap instrumen antara rater dan peneliti. Dengan demikian, instrumen yang dikembangkan dapat digunakan sebagai pedoman untuk menganalisis muatan penalaran deduktif dan induktif pada buku teks matematika.

5. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Ibu Dr. Dra. Raden Rosnawati, M.Si., selaku dosen pembimbing yang telah membimbing, memberikan arahan, dukungan, serta masukan yang sangat membangun dalam penyusunan penelitian ini.

6. REKOMENDASI

Saran serta rekomendasi untuk penelitian selanjutnya agar perlu mengkaji secara lebih mendalam untuk memperjelas karakteristik, indikator, dan contoh penalaran deduktif serta induktif dalam instrumen agar interpretasi antar-rater lebih seragam khususnya pada penyajiannya di buku teks matematika.

7. REFERENSI

- Akker, J. V. (2010). Building Bridges: How Research May Improve Curriculum Policies and Classroom Practices. *CIDREE*, 175-195.
- Ayalon, M., & Even, R. (2010). Mathematics Educators Views on the Role of Mathematics Learning in Developing Deductive Reasoning. *International Journal of Science and Mathematics Education*, 8, 1131-1154.
- Boateng, G., Neilands, T., Frongillo, E., Melgar-Quinonez, H., & L. Young^{1,5}, S. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Frontiers in Public Health*, 6, 1-18.
- Cole, R. (2024). Inter-Rater Reliability Methods in Qualitative Case Study Research. *Sociological Methods & Research*, 53 (4), 1944-1975.
- De Vet, H., B. Terwee, C., Mokkink, L., & L. Knol, D. (2011). Chapter 5 : Reliability. In H. C. De Vet, C. B. B. Terwee, L. B. Mokkink, & D. L. L. Knol, *Practical Guides to Biostatistic and Epidemiology Measurement in Medicine* (pp. 96-149). Cambridge: Cambridge University Press.
- Dockx, J., Bellens, K., & Fraine, B. D. (2020). Do Textbooks Matter for Reading Comprehension? A Study in Flemish Primary Education. *Frontiers in Psychology*, 1-19.

- Harel, G., & Weber, K. (2020). Deductive Reasoning in Mathematics Education. *Springer, Cham : Encyclopedia of Mathematics Education*, 183-189.
- Jeanotte, D., & Kieran, C. (2017). A Conceptual Model of Mathematics Reasoning for School Mathematics. *Educ Stud Math* 96, 1-16.
- Jia, W., & Zhang, P. (2023). Rater Cognitive Processes in Integrated Writing Tasks: from the Perspective of Problem-Solving. *Language Testing in Asia*, 1-20.
- Krippendorff, K. (2004). *Content Analysis: an Introduction to Its Methodology*. United States of America: Sage Publications.
- Lamprianou, I., Tsagari, D., & Kyriakou, N. (2023). Experienced but Detached from Reality: Theorizing and Operationalizing the Relationship Between Experience and Rater Effects. *Assessing Writing*, 56.
- Marx, R., Menezes, A., Horovitz, L., Jones, E., & Warren, R. (2003). A Comparison of Two Time Intervals for Test-Retest Reliability of Health Status Instruments. *Journal of Clinical Epidemiology*, 730-735.
- Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha Calculator–Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX*, 12.
- Moguel, L. E., Landa, E. A., & Cabanas-Sanchez, G. (2019). Characterization of Inductive Reasoning in Middle School Mathematics Teachers in a Generalization Task. *International Electronic Journal of Mathematics Education*, 14 (3), 563-581.
- Morris, A. (2002). Mathematical Reasoning: Adults' Ability to Make the Inductive-Deductive Distinction. *Cognition and Instruction*, 79-118.
- NCTM. (2004). *CSSU Math Framework* . United States of America: The National Council of Teachers of Mathematics.
- Nguyen, R., Brooks, M., Bruno, R., & Peacock, A. (2018). Behavioral Measures of State Impulsivity and Their Psychometric Properties: A Systematic Review. *Personality and Individual Differences*, 67-79.
- OECD. (2018). *PISA 2022 Mathematics Framework (Draft)*. Paris: OECD Publishing.
- Otten, S., & Gilbertson, N. J. (2014). The Mathematical Nature of Reasoning and-Proving Opportunities in Geometry Textbooks. *Mathematical Thinking and Learning*, 16(1), 51-79.
- PSKP Kemendikbud. (2022). *Risalah Kebijakan Nomor 11, Oktober 2022 Mengatasi Masalah Penyediaan Buku Teks Kurikulum Merdeka*. Jakarta: Pusat Standar dan Kebijakan Kurikulum (PSKP).
- Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, & Psikometrian)*. Yogyakarta: Parama Publishing.
- Sukendra, I. K., & Atmaja, I. K. (2020). *Instrumen Penelitian*. Bali: Mahameru Press.

- Thompson, D. R., Senk, S. L., & John, G. J. (2012). Opportunities to Learn Reasoning and Proof in High School Mathematics Textbooks. *Journal for Research in Mathematics Education*, 43(3), 253-295.
- Ulfah, A. A., Kartono, & Susilaningsih, E. (2020). Validity of Content and Reliability of Inter-Rater Instruments Assessing Ability of Problem Solving. *Journal of Educational Research and Evaluation*, 1, 1-7.
- Utari, T., & Hartono. (2019). Muatan Penalaran dan Pembuktian Matematis Pada Buku teks SMA Kelas X Kurikulum 2013. *Jurnal Riset Pendidikan Matematika*, 6(1), 1-13.
- Valverde, G. A., Bianchi, L., Wolfe, R., Schmidt, W., & Houang, R. (2002). *According to the Book Using TIMSS to Investigate the Translation of Policy Into Practice Through the World of Textbooks*. New York: Springer Science+Business Media New York.
- Wyse, A. (2021). How Days Between Tests Impacts Alternate Forms Reliability in Computerized Adaptive Tests. *Educational and Psychological Measurement*, 644-667.
- Zhang, D., & Qi, C. (2019). Reasoning and Proof in Eight-Grade Mathematics Textbooks in China. *International Journal Educational Research*, 98, 77-90.