

A ZSÚRIZÉS MINT ÉRTÉKELÉSI MÓDSZER A VIZUÁLIS NEVELÉSBN

Kárpáti Andrea, Zempléni András

Eötvös Loránd Tudományegyetem, Neveléstudományi Tanszék

Norman D. Verhelst, Niels H. Velduijzen és Diederik W. Schönau

CITO, Arnhem, Hollandia

Bevezetés

1993-ban a Művelődési és Közoktatási Minisztérium megbízásából, a holland-magyar államközi kulturális csereegyezmény keretében az új érettségi rendszer kialakítását segítő kutatás indult a holland Értékelési és Vizsgaközpont (CITO), az Országos Közoktatási Intézet és a Magyar Iparművészeti Főiskola részvételével. Az Új Pedagógiai Szemle 1995/8. száma bemutatja a CITO-t és a matematika és biológia tantárgyak magyar viszonyokra alkalmazott vizsgáztatási módszereit, melyeknek kidolgozását *Mátrai Zsuzsa* koordinálta és a *Horváth Zsuzsa* által vezetett, a holland anyanyelv és irodalom tantárgyak vizsgáztatási módszereit a magyar irodalom vizsga számára adaptáló kísérlet eredményeit. A kutatásban a matematika, angol nyelv, anyanyelv, biológia és rajz tantárgyakat vonták be. A már működő Rajzi Érettségi Munkacsoport tagjait kértük fel, vegyenek részt a rajzi vizsgarendszer adaptálásában. A holland rajzi érettségi leírását lásd *Schönau* (1994). A holland-magyar érettségi projektről számol be *Kárpáti* (1995), *Kárpáti* (1996), *Schönau* és *Kárpáti* (1996) valamint *Kárpáti* és *Schönau* (megjelenés alatt).

A CITO rajzi érettségi munkacsoportja, amely 18 éve készít feladatsorokat rajz – művészettörténet, textilművészet, környezetkultúra (kézművesség, design és iparművészeti tervezés) tantárgyak számára, 1981 óta kapcsolatban van vizuális képességkutató csoportunkkal. Az elméleti és gyakorlati feladatokat egyaránt tartalmazó vizsgát a nemzetközi rajzpedagógia a világ leghatékonyabb ilyen rendszereként tartja számon.

Hollandiában a vizuális nevelés körébe tartozó *három tantárgyból* lehet vizsgázni:

- a) Kétdimenziós alkotás (festés, grafika, fotó);
- b) Három dimenziós alkotás (szobrászat, kerámia, tárgykészítés);
- c) Textilművészet (nyomott- és szövöttanyag-tervezés, viselettervezés és -készítés, hurkoló technikák).

Mindhárom tantárgyból van *írásbeli teszt, gyakorlati feladat és szóbeli érettségi*. A központi ábrázoló feladatok tantárgyanként 6 témát kínálnak fel, amelyekből a tanulók egyet választanak ki és *projekt*-formában, január végétől áprilisig, összesen 28 tanóraban,

az iskolában, szaktanári felügyelet alatt, de segítség nélkül, önállóan oldanak meg. A gyakorlati feladatokat és ezek értékelési utasításait is a CITO szakértői dolgozzák ki és küldik meg az iskoláknak. Az ábrázoló vizsga tehát projekt-feladat, amelynek témalistáját a diákok januárban kapják kézhez. A javítást egy iskolakörzet rajztanárai közösen végzik, központi előírások alapján. A két bíráló által adott jegyek átlaga lesz a javasolt osztályzat. A munkákat kerületi értékelő központokban újra megvizsgálják, esetleg korrigálják az értékelést. A gyakorlati feladatok eredményeit kiállítás és munkanapló formájában mutatják be a tanulók. Nyilvánvaló, hogy a vizsga megbízhatósága attól függ, mennyire sikerül megbízható és a zsűrorok által következetesen használt *értékelési szempont-sort* készíteni a projekt-feladatok megítélésére. Ebben az írásban azt a kísérletet mutatjuk be, amellyel a holland típusú érettségit kipróbáló magyar csoport a *gyakorlati vizsga zsűrizési értékelésének megbízhatóságát* kívánta felmérni és javítani.¹

1994 februárjában kiadtuk a 12 holland projekt-feladatot és 21 magyar iskola 207 tanulója a hollandokkal egyidőben, azonos tartalmú gyakorlati vizsgát tett. Három hónapig önállóan, de felügyelet alatt megoldottak egyet a hat kétdimenziós (festészet, grafika, tervezés, fotó) vagy a 6, három dimenziós (szobrászat, installáció, tárgykészítés) feladatból. (A feladatokat – akárcsak a holland diákok – szabadon választották.) A Rajzi Érettségi Munkacsoport az 1994/95-ös tanévben tovább bővült: 36 tanár több, mint 500 tanuló kísérleti vizsgáztatását végezte el. A III. és IV. osztályos tanulók 1995 januárja és júniusa között mind a gyakorlati feladatokat, mind a tesztet megoldották és a hollandokkal azonos formájú és tartalmú vizsgát tettek. A vizsga-projektet a tanulók saját rajztanára és egy megbízott szakértőnk értékelte.²

A zsűri kísérlethez az 1994-es vizsga anyagából választottunk ki egy expresszív és egy design feladatot. A feladatokat projekt-szerűen, tehát háttér-kutatásokat végezve, *a terveket, vázlatokat, variációkat és az ötletek esetleges művészettörténeti forrását* is bemutatva, kiállítás-ként kellett elrendezni. A zsűrorok így nem csak a kész alkotást, hanem a vizuális gondolkodás fejlettségét tükröző valamennyi munkafázist bírálták el. (A tanulók vizsga-munkáinak gyűjteményét a továbbiakban *kollekciónak* nevezzük.) Az expresszív feladat „*Pillantás a végtelenbe*” címmel festmény vagy grafika készítését kérte, a design feladatban „*Tervezz tapéta-mintákat tengerparti szállodába!*” címmel egy szálloda társalgójába, a gyermekszobába és egy fürdőszobába kellett tapétát tervezni, a tengerpartot jellemző motívumok felhasználásával. Az utasítás sem a technikára, sem a méretre vagy az elkészítendő kiállítási anyag részeire vonatkozólag nem tartalmazott további megkötéseket. Ez a két téma a magyar kísérleti vizsgán a legnépszerűbbek közé tartozott. Budapesten, 1995 júniusában 31 magyar és hat külföldi szakértő közreműködésével az adaptált holland értékelési kritérium-rendszerrel (ld. Függelék) zsűriztük az alkotásokat. Az 1. feladatból 30, a 2., design feladatból 27 munkát választottunk ki a zsűri-kísérlet-

¹ A Rajzi Érettségi Munkacsoport megbízása nem az érettségi korszerűsítésére szőtt, csupán a holland tapasztalatok megismerésével és a projekt-módszer valamint az egy téma köré csoportosuló művészettörténeti-műkritikai teszt adaptálásával és kipróbálásával a vizuális nevelési vizsgarendszer megújítására kellett ajánlásokat készíteni. Javaslatunkban három vizuális érettségi tantárgy szerepel: az *ábrázolás* (grafika, festészet és a geometrikus leképező rendszerek, az ábrázoló geometria alapjai) a *művészettörténet* és a *környezetkultúra*. Vizsgafeladat-bankunkat 1993 februárjától a két előbbi tantárgyból építettük fel.

² A tanulók feladat-választásainak összegzését és a választások elemzését ld.: Kárpáti (1995).

hez. Mivel arra törekedtünk, hogy a munkák jól reprezentálják a kísérlet során született projekt alkotások átlagos színvonalát, gyakorlatilag az összes, a témákban készült munkát felvettük a zsűrizendők közé. Csak azt a két feladatsort hagytuk ki, amelyeknek egy-egy része a kiállítások, bemutatók során elveszett. Ennél az első zsűri-kísérletnél a zsűrorok a szempont-listát követve, valamennyi szempont alapján, egymás után bírálták el az alkotásokat. 1996 tavaszán megismételtük a zsűri-kísérletet, azonos szempont-rendszerrel. A második zsűrizéskor az értékelők egy-egy szempont szerint pontozták végig valamennyi kollekción, tehát lehetőségük volt az adott szempont szerinti teljesítmények összevetésére, a kollekciónak egymáshoz való viszonyítására. Valamennyi zsűritagunk gyakorló rajztanár volt, körülbelül fele-fele arányban budapesti és vidéki iskolák oktatói. Az értékelést a szempontok megbeszélése (két felkészítő tréning) előzte meg. A bírálati munka tapasztalatairól videóra rögzített megbeszélésen számoltak be a résztvevők.

Ebben a tanulmányban a zsűrizés, mint értékelési módszer megbízhatóságát elemezzük a nemzetközi projekt-zsűri tapasztalatai alapján. A zsűrorok által használt értékelési szempontrendszert a CITO rajzi vizsgáinak tízéves tapasztalatai és a két tanéves magyar iskolakísérlet alapján *Diederik Schönau és Kárpáti Andrea* állította össze. (Az értékelő szempontrendszert a Függelékben közöljük.) Írásunk végén összehasonlítjuk az 1995-ös adatokat ugyanazon alkotások 1996-ban megismételt zsűrizése során kapott adatokkal. Az ismételt kísérlet célja az volt, hogy megvizsgáljuk, képesek-e objektíven értékelni a projekt-munkákat olyan zsűrorok, akik nem vettek részt a kutatásban, tehát az értékelési szempontokon túl nem rendelkeznek egyéb információkkal az alkotások készülése körülményeiről. Ez a második kísérlet az általunk javasolt érettségi vizsgát modellezte: bemutatta, mennyire képesek árnyaltan, megbízhatóan értékelni külső, az iskolában nem tanító pedagógusok a tanulók projekt alkotásait.

Az elemzés során a zsűritagok által adott pontszámok összességével dolgoztunk (ez több, mint 6000 adat). Ennek az eredeti adatbázisnak a használatát alátámasztotta a CITO matematikus szakembereinek elemzése is, amely szerint, ha a szempontokra adott válaszokat kvalitatív mennyiségeknek tekintjük és ezután homogenitás-vizsgálattal számszerűsítjük a válaszokat, akkor – homogén zsűriket feltételezve – a kapott érték nem tér el lényegesen az eredeti pontszámok számértékétől.

Homogenitás-vizsgálat

Elméleti alapok

A zsűrorok szavakkal megfogalmazott állítások alapján bírálták el a kollekciónkat, és a zsűritagok által adott értékelések is szavakkal megfogalmazott minősítéseknek feleltek meg. Bár remélhető, hogy ezek a számok valamely mennyiség kifejezői is, az elemzés szempontjából ezeket minőségi címkéknek tekintettük, azaz a *kritériumokat minőségi változókként kezeltük*. A homogenitás-vizsgálat célja, hogy az eredeti értékeléseket mint címkéket zsűritagról zsűritagra számokkal helyettesítsük – ezeket pontszámoknak nevezzük. Az egyes kollekciónak végső pontszáma az egyes zsűritagoktól ily módon kapott össz-

pontszámok átlaga. A számokat úgy választjuk, hogy minden egyes kollekciónak végső pontszáma – amennyire csak lehet – emlékeztessen a zsűritagok által meghatározott értékeléshalmazra. Formálisabban meghatározva, ez azt jelenti, hogy a következő két feltevéssel élünk:

- Az i -edik kollekciónak minősége vagy jósága egy x_i -vel jelölt számmal adható meg;
- Ha a j -edik zsűritag az m címkét adja (egy bizonyos kritériumnál) két különböző kollekciónak, akkor feltételezzük, hogy az ő véleménye szerint ez a két kollekciónak az adott kritérium szerint hasonló. Ezért azt feltételezzük, hogy a „ j -edik zsűritag az m címkét adja az i -edik kollekciónak” esemény egy y_{jm} számmal írható le.

A homogenitás-elemzés az x_i és y_{jm} számok megtalálásának módszere. Durván szólva, ez a módszer azt feltételezi, hogy ha a j -edik zsűritag az m címkét adja az i -edik kollekciónak, akkor x_i és y_{jm} nem különbözhet túlságosan. Az, hogy a fenti módszerrel megtalált optimális megoldás nem használható a jövőben, a zsűritagok különbözőségén múlik. A módszerünk szerint abból, hogy egy bizonyos kritérium valamely kategóriáját zsűritagonként számszerűsítjük, nem következik, hogy az adott kategória ugyanazzal a pontszámmal írható le, ha egy másik zsűritagról van szó. Azaz, a számszerűsítés személyes, és így a jövőben minden zsűritagnak viselnie kellene saját számszerűsítését. Egy, a jövőbeli értékeléseknél is használható számszerűsítést kaphatunk az úgynevezett *korlátozottan optimális pontszámok* révén. Ekkor az a követelmény, hogy a számszerűsítés azonos legyen minden zsűritagra.

A megfigyelt pontszámok, melyek egy spontán megközelítésből adódnak, olyan skálát eredményeznek, mely rosszabb minőségű az optimális vagy korlátozottan optimális pontszámoknál. De mindez nem jelenti azt, hogy az ezeken a pontszámokon alapuló elemzés értéktelen lenne, sőt, ahogy ez az eredményeket leíró részből kiderül, az eredeti pontszámok majdnem olyan jók, mint a korlátozottan optimálisak. Ezt az állítást a korrelációk vizsgálatával támaszthatjuk alá (lásd az 1. táblázatot).

1. táblázat. Az optimális (x), a korlátozottan optimális (x^*) és a megfigyelt pontszámok (megf) korrelációja kritériumonként

Kritérium	Felism	Terv	Művkif	Skicc	Választ	Kül	Expr	Anyag	Techn	Konc
x, x^*	0,91	0,95	0,90	0,94	0,92	0,94	0,92	0,91	0,95	0,95
$x, \text{megf.}$	0,92	0,94	0,91	0,90	0,90	0,94	0,92	0,92	0,93	0,92
$x^*, \text{megf.}$	0,99	1,00	1,00	0,91	0,99	0,98	1,00	1,00	0,98	0,97

A kritériumok listája és a táblázatokban használt rövidítésük:

- A választott téma felismerhetősége (Felism)
- A tervezőmunka, a diák szándékainak felismerhetősége (Terv)
- A választott művészi kifejezőmód originalitása (Művkif)
- A tanuló egy sor különféle megoldással kísérletezett, amelyekről skiccek és változatok tanúskodnak (Skicc)
- A tanuló választása a megoldási lehetőségek/tervek/változatok között (Választ)

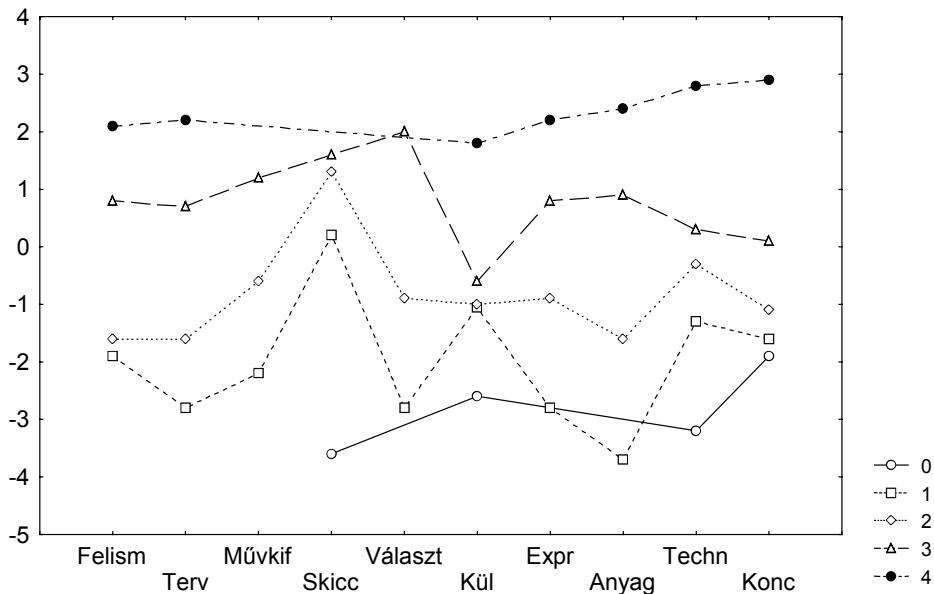
- Különleges képi hatások, amelyek a feladatokhoz kapcsolódnak (Kül)
- A többi vizuális kifejezőeszköz, amelyet a tanuló kiválasztott (szín, kompozíció, ritmus stb.) kapcsolata a tanuló alkotói szándékával (Expr)
- Az anyag/technika megválasztásának sikeressége a tanuló alkotói szándékainak megvalósítása szempontjából (Anyag)
- Technikai tudás, ábrázoló képesség (Techn)
- A téma-orientáltság szintje, a tanuló koncentrációja (Konc)

Az értékelési kritériumokat és kategóriákat részletesen ismertetjük a Függelékben.

A korrelációk igen magasak a korlátozottan optimális és a megfigyelt pontszámok között (az egyetlen lehetséges kivétel a vázlatkészítésre illetve a vázlatok minőségére vonatkozó kritérium, rövidítve: „Skicc”). Ez azt jelenti, hogy a megfigyelt és a korlátozottan optimális pontszámok lényegében ugyanazt az eredményt adják. Ugyanakkor az optimális pontszámoknak és a korlátozottan optimális illetve a megfigyelt pontszámoknak a korrelációja, bár meglehetősen nagy, lényegesen és szisztematikusan alacsonyabb. A különbségek a zsűritagok közötti szisztematikus eltérésekből adódnak.

A számszerű értékeként kezelt eredeti kategóriák legfontosabb tulajdonsága a szomszédos értékek azonos távolsága, vagyis az az elvárás, hogy az egyes kritériumokra adott pontszámok szomszédos tagjai között azonos érték-távolság legyen. A homogenitás-vizsgálat korlátozottan optimális pontszámai mutatják, hogy mivel a megfigyelésekből minél több információt akarunk egyetlen pontszámra sűríteni, ez a feltételezés nem bizonyul megalapozottnak. Például az 1. kritériumnál (témafelismerés) az 1. ábrából világosan látható, hogy az 1. és 2. kategória (a téma felismerhetetlen ill. hosszabb ideig tartó megfigyelés alapján felismerhető, de nem meggyőzően) szinte megkülönböztethetetlen. Ugyanez a helyzet a 6. kritériumnál (ez függött a feladattól) és a 10. kritériumnál (koncentráció) is. Valószínű, hogy ezen kritériumok megfogalmazásánál pontosabban kell érzékeltetni a különbséget az egyes kategóriák között. A korlátozottan optimális pontszámnál is megőrződik az azonos távolság a 3., 7., és a 8. kritériumnál (művészi megoldás, vizuális kifejezés ill. anyaghasználat) és – valamelyest kevésbé – a 2. kritériumnál (tervezés).

A legjelentősebb eltérés a 4. kritériumnál (skiccek, vázlatok készítése) figyelhető meg, ahol még a megfigyelt és korlátozottan optimális pontszámok közötti monoton kapcsolat is elveszett. Egy másik váratlan eredmény ezzel a kritériummal kapcsolatosan az, hogy a 0 kategória korlátozottan optimális pontszáma kiugróan alacsony. Ez az érték olyan messze van az összes többi korlátozottan optimális pontszámtól, hogy az a benyomás alakul ki, hogy a 4. kritériummal kapcsolatosan a legfontosabb értékelés így is megfogalmazható: „0 (nincsenek skiccek)” vagy „nem 0 – vannak skiccek”. Azaz a 4. kritériumot a zsűritagok többé-kevésbé bináris kritériumként kezelik. A 10. kritérium (koncentráció) esetén is hasonló – csak fordított – jelenség lép fel: ott a 4. kategória áll szemben az összes többivel. Ezek az interpretációk természetesen nem bizonyító erejűek: csak segítséget jelentenek abban, hogyan rendeljünk értékeket az egyes kategóriákhoz, vagy esetleg abban, hogyan fogalmazzuk át a jövőben néhány kategória megszövegezését.



1. ábra

A „Pillantás a végtelenbe” téma kritériumonkénti korlátozottan optimális pontszámai

Az egyes kritériumok kevésbé homogén értékelése kétféle problémára vezethető vissza: egyrészt elképzelhető, hogy a kritérium – vagy a kategória – rossz (rosszul definiált, értelmetlen vagy az adott alkotásokra nem alkalmazható), de az is lehetséges, hogy a zsűri nem homogén, egyes tagjai túlzottan szigorúak, míg mások igen engedékenyek. Mindkét feltételezést vizsgálni fogjuk és e tanulmány 4. részében még visszatérünk egy könnyen interpretálható érdekes modellre, ahol a zsűritagok közötti eltéréseket lineáris függvénnyel írjuk le.

A megfigyelt pontszámok elemzése

Itemanalízis

A zsűrizéses értékelés kulcskérdése, hogy mennyire használható a munkák elbírálására szolgáló értékelési kritérium-rendszer. Az itt alkalmazott pontozási útmutató jóságát mutatja, hogy a 2. táblázatból leolvasható megbízhatóságok meglehetősen magasak, de szembetűnő különbségek figyelhetők meg a kritériumok között. A vázlatokat minősítő „Skicc” és a témaspecifikus („Kül” rövidítésű, a tervezési illetve expresszív megoldást kívánó) kritérium adja a legalacsonyabb értékeket, sugallva, hogy ez a két kritérium kissé „kilóg” a többi közül. Ez felkészítési problémákat vet fel, hiszen az itemek megfogalma-

zása teljesen egyértelmű. A „Skicc” kritérium a vázlatkészítés színvonalát minősíti. A témaspecifikus kritériumok a „Pillantás a végtelenbe” témánál a perspektíva alkalmazását, a design feladatnál a forma és a funkció kapcsolatát (a szoba rendeltetése és a tapéta szín- és formavilága közötti összefüggést) értékelik. *A zsűriző rajztanárok különbözőképpen ítélik meg, mennyire sikeres egy műben egy jól definiálható stílusjegy megjelenése.* Nehezebb értékelési feladat ez számukra, mint az esztétikai minőség, az eredetiség – a laikusok számára értékelhetetlennek tűnő – minőségének megítélése. Az experimentális esztétika tapasztalatait figyelembe véve azonban az eredmény egyáltalán nem meglepő: azonos képzettségű, foglalkozású, társadalmi helyzetű és korú zsűrörök hajlamosak azonos vizuális minőségeket szépnek látni. A stílusjegyek felismerése azonban hosszabb tanulást, alaposabb tréninget igényel, mint amire ebben a vizsgálatban lehetőségünk volt. A vizsgamódszer alkalmazásakor azonban – mint az eredmények mutatják – nem fordulhat elő, hogy a tanárok alapos felkészülés nélkül vállalkozzanak a projekt-munkák megítélésére. *Úgy tűnik, a mindennapi értékelő-osztályozó munka nem készít fel a többszempontú, a szempontokat egyenrangúan figyelembe vevő projekt-értékelésre.*

2. táblázat. A kritériumok mint itemek analízise

Kritérium	Max	Átlag	P(%)	Sd.	R _{it}	R _{ir}
Összbe	10	5,64	56	2,80	84	74
Felism	4	2,71	68	1,10	72	66
Terv	4	2,66	66	1,12	73	67
Művkif	4	2,28	57	1,17	80	76
Skicc	3	1,36	45	0,91	62	56
Választ	3	2,05	68	0,82	76	73
Kül	4	2,02	50	1,47	67	59
Expr	4	2,39	60	1,11	78	74
Anyag	4	2,43	61	1,11	72	66
Techn	4	2,16	54	1,26	82	78
Konc	4	2,24	56	1,24	86	82

Elérhető maximális pontszám : 48

Átlagos pontszám: 27,93

A minták száma: 763

Reliabilitás (f): 0,90

Szórás (Sd): 10,85

Átl. P-érték: 58,20

Standard hiba: 3,35

P%: az átlagpontszám a maximum %-ában, Sd: az adott kritérium szórása, R_{it} ill. R_{ir}: az item és az összpontszám korrelációja – beleértve illetve kihagyva az adott item pontszámát.

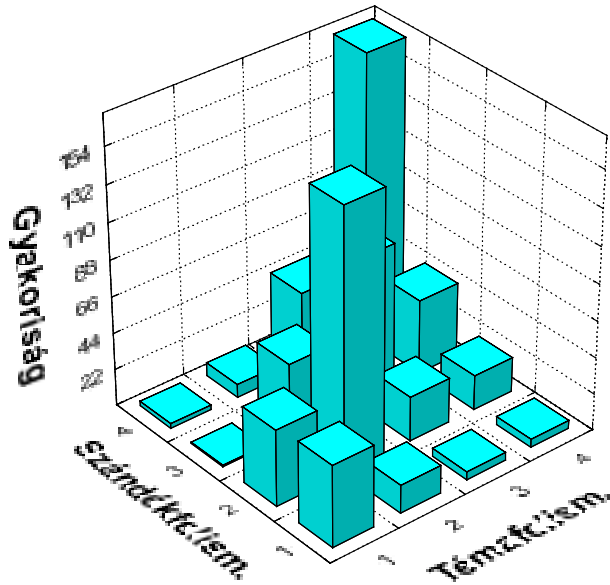
A kritériumok és a zsűritagok viselkedése

Fontos észrevétel volt az első zsűrizésnél, hogy a kollektívokkal kapcsolatban kialakított előzetes vélemény, ízlésítélet nagy szerepet játszott az egyes szempontok szerinti értékelésnél is. Az egyes zsűritagokat valamennyi szempont szerinti értékelésnél erősen befolyásolta globális ítéletük. Ha a kollektív tetszett, minden kritériumra magas pontszámot adtak neki, illetve fordítva: a kedvezőtlen összbemutató alapján minden szempontnál alacsonyan tartották a pontszámokat. Éppen ezért döntöttünk úgy, hogy a megismételt vizsgálatnál nem kollektívként, hanem szempontonként kapnak értékelő lapot a zsűrirok. Ez a módszer minden további nélkül alkalmazható az érettségi vizsgákra is, ahol egy iskolában átlagosan csak 10–15 munkát kell elbírálni. Az egyes kritériumok összefüggésének vizsgálatát szemléletesen az alábbi három dimenziós hisztogramokkal végezhetjük.

Ezzel az elemzéssel feltárható, mennyire értették a zsűritagok az egyes kritériumok jelentését. A 2. ábrán az alkotói szándék és a központi tartalom, téma felismerhetőségének értékelése található. Az első a munkanapló alapján ítélt meg, hiszen ebben mondja el, vázlatokkal illusztrálva, első ötleteit és választott megoldásait a projekt készítője. Ha a szemlélő felismeri a kiválasztott témát és leolvassa a tartalmat (tehát nem egyszerűen tájképet lát, hanem a „Pillantás a végtelenbe” téma megjelenítését), magas pontszám adható. Az eredményekből azonban jól látható, hogy az esetek többségében ugyanazt a pontszámot kapták a kollektívok erre a két szempontra (a korrelációs együttható 0,7). Egynél nagyobb eltérés pedig alig-alig fordult elő. A rengeteg egyezés azt mutatja, hogy a zsűrirok pontszámaikat túl gyakran alapozzák az általános bemutatóra, nem tekintik át a terveket, kevés figyelmet fordítanak a munkanaplókra, amelyek az első ötletek születésétől rögzítik az alkotó folyamat állomásait. Fel sem tételezik, hogy a munka készítőjének szándékai egészen mások voltak, mint amit a végeredmény sugall. A magyar rajztanítás legfontosabb eltérése a hollandtól éppen az oktatás legfontosabb céljának megítélése. A magyar rajztervek centrumában a technikai készségek fejlesztése, a „rajzolás tanítása” áll, míg a holland rajztanárok a képi ötleteket, a tervezést, a variációk készítését ennél sokkal fontosabbnak tartják. Nem véletlen, hogy a magyar zsűriroknak problémát okoz egy olyan helyzet, ahol a tervek lényegesen jobbak, mint a végeredmény. *A tervek megítélése helyett szívesebben hagyatkoznak a kész műről kialakított bemutatóikra. Ez az eljárás teljesen rendjén való, ha képzőművészeti alkotásokat válogatunk egy kiállításra, de használhatatlan, ha átlagos képességű vizsgázók munkáit értékeljük azzal a céllal, hogy vizuális képességeikről (nem pedig tehetségükről) minél árnyaltabb képet alkossunk.* A projekt rendszerű vizsga bevezetésének alapvetően fontos feltétele, hogy ezt az attitűdöt megváltoztassuk.

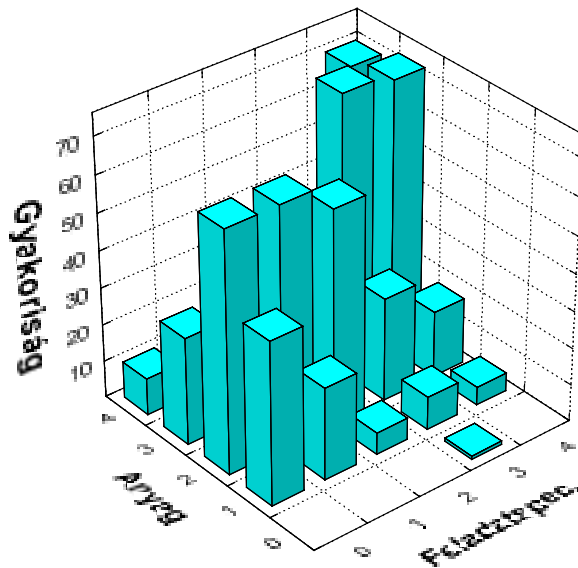
Számos kritérium esetében szerencsére árnyaltabb az értékelés, jobban különböznek a pontszámok. A 3. ábra egy olyan kritérium-párt mutat, ahol jobban megoszlának a pontszámok (a korrelációs együttható 0,5).

A zsűrizés mint értékelési módszer a vizuális nevelésben



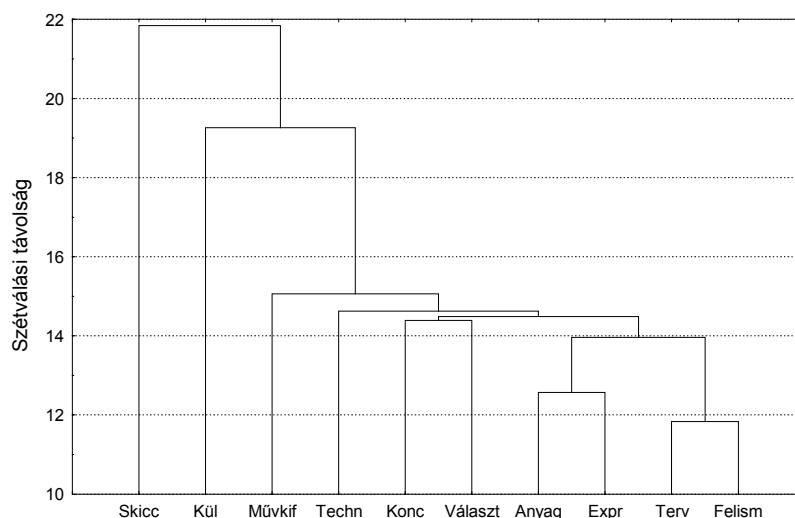
2. ábra

Az alkotói szándék („üzenet”) és a téma felismerhetőségének együttes ábrázolása



3. ábra

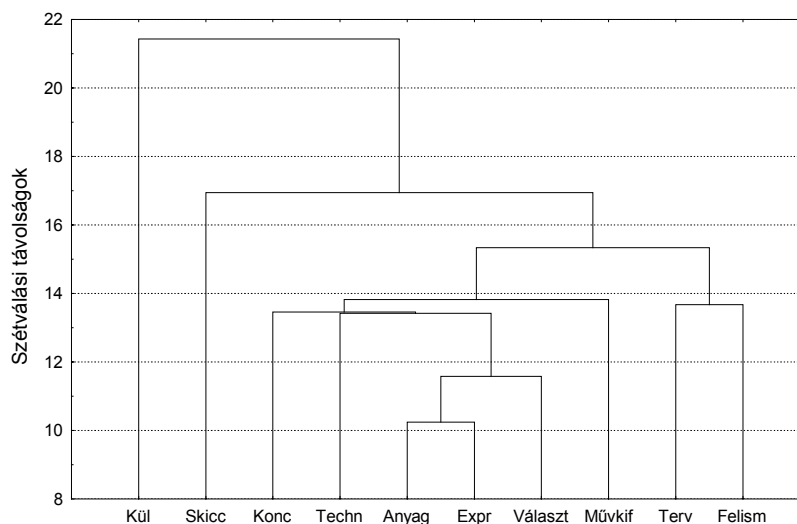
Az anyagok és a technikák megválasztása és a feladat-specifikus kritérium (perspektíva, illetve forma és funkció összeegyeztetése) együttes ábrázolása



4. ábra

Fa-diagram a tíz kritériumra, expresszív feladat

Az alacsonyan (kis szétválasztási távolságnál) elhelyezkedő vízszintes vonallal összekötött kritériumokra (kritérium-csoportokra) kapott pontszámok hasonlóak egymáshoz



5. ábra.

Fa-diagram a tíz kritériumra, design feladat

Az alacsonyan (kis szétválasztási távolságnál) elhelyezkedő vízszintes vonallal összekötött kritériumokra (kritérium-csoportokra) kapott pontszámok hasonlóak egymáshoz

A többi kritérium pontszámainak megoszlása lényegében e kettő között helyezkedik el, csupán a tervek, vázlatok minőségét értékelő kritérium mutat ennél gyengébb összefüggést a többi kritériummal.

Ha összehasonlítjuk az expresszív (képzőművészeti jellegű) és a design feladatot, látható, hogy a kritériumok viselkedése a két feladatnál nagyon hasonló. Mindkét esetben *az anyag és a technika megválasztását és a vizuális kifejezőeszközök használatát minősítő kritérium helyezkedik el egymáshoz legközelebb (vö. 4. és az 5. ábrát)*. A zsűrorok egy-egy projekt-munka esetében nem tesznek különbséget a kétféle teljesítmény között. Szinte mindig közel azonos pontszámmal értékelnek egy-egy tervezői és egy manuális tevékenységet: a médiumválasztást (az ismert illetve rendelkezésre álló anyagok és technikák közötti döntést) és ezek használatát, a technikai tudást vagy ügyességet, jártasságot. Ezt a felismerést megerősíti a 6. ábrán látható vizuális ábrázolás is.

Mindkét feladatnál probléma, hogy a jó terv csak akkor kap magas pontértéket, ha a megvalósítás is sikeres. Ha azt akarjuk, hogy a rajzi vizsga ne csak a leendő professzionális rajzoló teljesítményének értékelésére legyen alkalmas, a bírálókat meg kell tanítanunk a tervezés lépéseinek értékelésére, a jó terv felismerésére, a kevésbé sikeres (technikailag megoldatlan, ügyetlen) műben. Évszázados hagyományokat kell ehhez megváltoztatnunk: a legkevésbé formalista képzőművészeti irányzatok korában ugyanis éppen úgy a rajzkészség (a bravúrok utánzása) volt az akadémiák felvételi mércéje, mint azokban a századokban, mikor a realista leképezési módok még a korstílus műveléséhez elengedhetetlenek voltak. A konceptuális irányzatok korában azonban itt az idő – és a projekt-jellegű vizsga kínálta lehetőség – az árnyaltabb, a képi gondolkodást is méltányoló értékelésre. Hogy ez milyen nehéz dolog, azt zsűri-kísérletünk is bizonyítja.

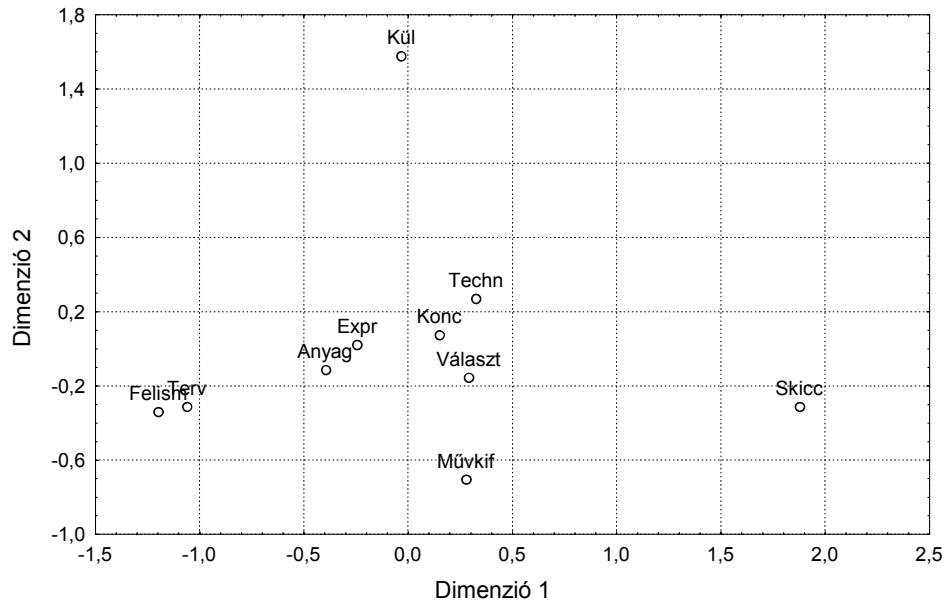
Az itemek elemzését fejezzük be annak vizsgálatával, hogy mennyire közelíthető meg a globális értékelésre kapott pontszám a kritériumokra kapott pontszámok lineáris függvénye segítségével (3. táblázat).

3. táblázat. Az összpontszám közelítése: többszörös regresszió – a négy legfontosabb változó együtthatója

	Súly
A téma felismerhetősége (Felism)	0,437
A művészi kifejezés originalitása (Művkif)	0,780
Vizuális kifejezőeszközök használata (Expr)	0,598
Technikai tudás (Techn)	0,524
Konstans	0,040

A lineáris modell illesztésénél viszonylag magas R értékek adódtak (0,6–0,7), ezek mutatnak némi összefüggést, de ahhoz nem elég nagyok, hogy bizonyítsanak egy egyszerű lineáris kapcsolatot. Esetünkben ez az eredmény igen fontos, hiszen bizonyítja, *hogy a képzőművészeti alkotások zsűrizésénél használatos, „globális” megítélés kevésbé hatékony, mint a kritériumok szerinti részletes értékelés*. A „rápillantásos” módszer, mely a

mű által a szemlélőben keltett összbenyomást (sok esetben az igen változékony első benyomást) teszi a kiválóság mércéjévé, nem ad azonos eredményt, mint az alapos, részletekre kiterjedő, mérlegelő vizsgálat.



6. ábra

A kritériumok kétdimenziós skálázása (együtt a két feladatra)

A két feladat-típus összehasonlítása

Vizsgánk némileg hasonlít a képző- és iparművészeti zsűri munkájára, ahol a műtárgykollekciókat szakértők csoportja minősíti. A képző- és iparművészet rangjáról, az „alkalmazott művészetek” státusáról sok szó esik, de tudomásunk szerint senki sem vizsgálta még meg, *ugyanolyan hiteles és összemérhető-e egy képzőművészeti és egy iparművészeti zsűri munkája*. Vajon azonos nehézségű feladat-e az autonóm műalkotások és a meghatározott funkciót betöltő tárgyak megítélése? Az egyes értékelési kritériumokra adott pontszámok korreláció-elemzése alapján megállapítható, hogy mindkét feladat esetében erős összefüggés van az egyes szempontok között, tehát a kritériumok homogén csoportot alkotnak. Az viszont érdekes, hogy a „Pillantás a végtelenbe” című expresszív (képzőművészeti jellegű) feladatnál általában alacsonyabbak a korreláció értékek, mint a „Tervezz tapétát tengerparti szálloda nappalijába, gyermek játszószobájába és egy fürdőszobába!” című design feladatnál. Ez arra utal, hogy minden bizonnyal nem azonos megbízhatósággal értékelhetők a nyitott, kreatív feladatok és az alkalmazott művészeti, tervezési feladatok. Az összefüggés ezen kívül azt is mutatja, hogy a feladatok nem egészen egyformák matematikai szempontból sem. Az egyes szempontok statisztikai vizsgálatá-

ból kitűnt, hogy minden kritériumnál – sokszor szignifikánsan – alacsonyabbak az átlagos pontszámok a design témánál, mint az expresszív, az egyéni fantáziára, szabad képi kifejezésre építő feladatnál.

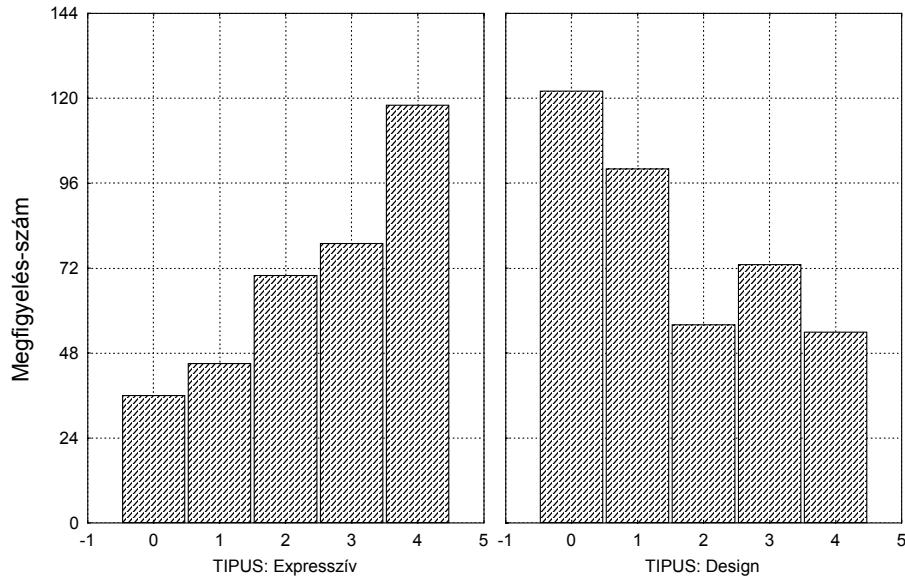
4. táblázat. A kritériumokra kapott pontszámok összehasonlítása a két feladatnál

	<i>z</i> -érték	<i>p</i> -érték
Összbenyomás	4,0870	0,00004
Felismerhetőség (Felism)	– 0,4868	0,62643
Tervezőmunka (Terv)	1,4068	0,15951
Művészeti kifejezőmód (Művkif)	2,8377	0,00455
Kísérletezés (Skicc)	– 1,7176	0,08589
Választás (Választ)	3,1584	0,00159
Különleges képi hatások (Kül)	8,8766	0,00000
Vizuális kifejezés (Expr)	4,1939	0,00003
Anyag-technika (Anyag)	2,8958	0,00378
Technikai tudás (Techn)	4,0350	0,00005
Téma-orientáltság (Konc)	2,6722	0,00754
Összpontszám	4,0128	0,00006

A *z*-érték a számolt rangstatisztika értéke. Ha azonos nehézségű lenne a két feladat, akkor *z* standard normális eloszlású lenne. A *p*-érték a szignifikancia-szint.

A zsűrizésre kiválasztott két munka alapvetően különböző alkotói habitust kívánt illetve engedett érvényesülni. Az egyik, a „Tervezz tapétát tengerparti szálloda helyiségeibe” című, a megrendelő feladatkiírására válaszoló, tehát részben kötött, *tervezési feladat* volt. A másik munka viszont a *szabad művészi kifejezést*, az egyéni stílust juttatta érvényre. Az 4. táblázatban a két feladat-típusra kapott pontszámokat hasonlítjuk össze. Az összehasonlítás az úgynevezett standardizált rangokon múlik, ami az egyes pontszámoknak a teljes, nagyság szerint rendezett mintában elfoglalt helye alapján számolható. Számos kritériumnál igen kis *p* érték található, ami azt mutatja, hogy ezeknél szignifikáns a különbség a két feladat pontszámai között (pozitív *z* érték esetén az expresszív feladat pontszámai, míg negatívnál a design feladat pontszámai a magasabbak – látható, hogy minden szignifikáns különbségnél az expresszív feladat kapott magasabb pontszámot; az „Összanal.” az analitikus kritériumok összege).

A legnagyobb különbséget a témaspecifikus kritérium („Kül”) mutatta. Ennek a hisztogramja látható a 7. ábrán.



7. ábra

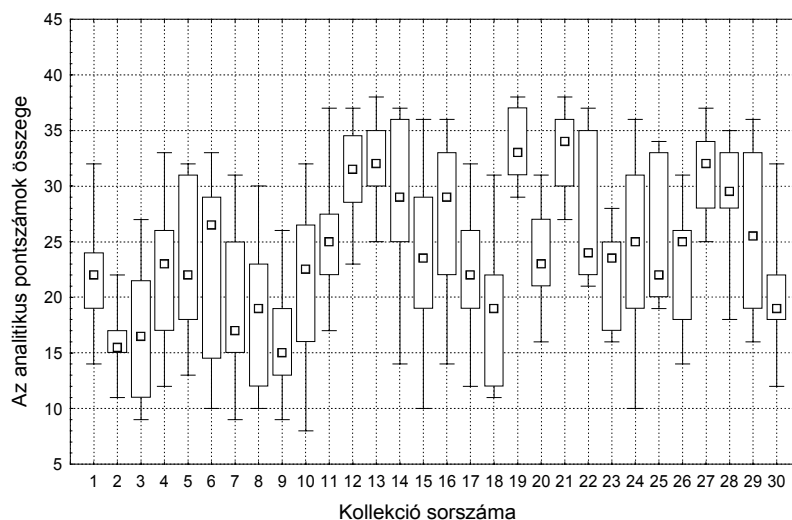
A témaspecifikus kritérium (kül): a perspektíva alkalmazása (az expresszív feladatnál) és a szobák funkciója (a design feladatnál) pontszámainak megoszlása

Az összpontszám megbízhatósága

Először vessünk egy pillantást az egyes kollektívok értékelésére. A 8. és 9. ábrán az egyes kollektívokra adott összpontszámok középértéke, a 25–75%-os határok és a teljes terjedelem látható. Ez az elemzés is igazolja, hogy a design feladat megbízhatóbban értékelhető, mint a képzőművészeti jellegű.

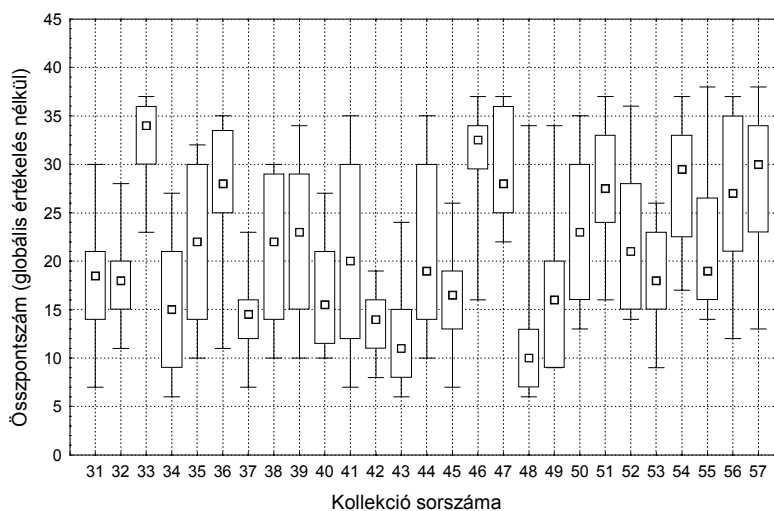
Néhány kiugró érték magyarázata az, hogy a zsűritag saját tanítványának művét bírálta. Őt olyan tanár is zsűror volt, akinek tanítványai résztvettek a kísérleti rajzi vizsgán is. Ez azonban nem magyaráz minden esetet, hiszen néhány más kollektív esetében előfordultak magas és alacsony összpontszámok egyaránt. Ez azt bizonyítja, hogy *vannak viszonylag egyértelműen megítélhető kollektívok és vannak vitát kiváltók is*. Ha projekt-módszerű vizsgát tervezünk, nélkülözhetetlenek a bemért „minta-kollektívok”, melyeket több szakértő minősített és amelyek az értékelési szempontokat illusztrálják.

A zsűrizés mint értékelési módszer a vizuális nevelésben



8. ábra

Az egyes kollekciók különböző zsűroroktól kapott összpontszámainak megoszlása (az ábrán a középvérték, a 25–75%-os határok és a teljes terjedelem látható; expresszív feladat)



9. ábra

Az egyes kollekciók különböző zsűroroktól kapott összpontszámainak megoszlása (az ábrán a középvérték, a 25–75%-os határok és a teljes terjedelem látható; design feladat)

A kritériumok kombinációjának megbízhatósága meglehetősen alacsony mind a megfigyelt értékek, mind a korlátozottan optimális pontszámok és a főkomponensek esetén is (egy javítást l. az 4. részben). Ha a megfigyelt változók átlagának megbízhatóságát és a két zsűritag esetét tekintjük példának (a megbízhatóság természetesen nő a zsűritagok számának növelésével), akkor a kombinált kritérium megbízhatósága 0,56, míg a külön-külön vett kritériumok megbízhatósága 0,35 és 0,54 között változik – a medián pedig 0,43 és 0,45 között. Ha viszont minden egyes kritériumot egy egykérdéses tesztnek tekintenénk, akkor a kritériumok kombinációja egy, az eredetihez képest 10-szer olyan hosszú teszt konstrukcióját jelentené. Ha a pontszámok függetlenek lennének, alkalmazhatnánk a Spearman-Brown formulát az összetett teszt megbízhatóságának becslésére, amely szerint $\rho_{10} = \frac{10\rho}{1+9\rho} = 0,89$ és ily módon a megfigyeltnél jóval magasabb megbízha-

tósági értékeket kapnánk. Mivel a talált értékek meglehetősen eltérnek ettől az előrejelzéstől, nem teljesül valamely feltétel, mely a Spearman-Brown formulához szükséges.

A teljes variancia három részre bontható: egy, a kollektcióknak tulajdonítható részre, egy, a zsűritagokhoz rendelhetőre és egy reziduális részre, amely a kollektció-zsűritag kölcsönhatást és a mérési hibát tartalmazza. A megfigyelt X_{ijc} értékeken és a kollektció valamint a zsűri hatások figyelembe vételén alapuló lineáris modell illesztése után fennmaradó reziduálisok korrelációi ugyanolyan nagyságrendűek, mint az eredeti pontszámok közötti korrelációk, azt mutatva, hogy fontos kölcsönhatások vannak jelen.

5. táblázat. Az interakció nélküli lineáris modell reziduálisainak korrelációi

	Felism	Terv	Művkif	Skicc	Választ	Kül	Expr	Anyag	Techn	Konc
Felism	1,00	0,73	0,34	0,32	0,46	0,55	0,46	0,50	0,40	0,51
Terv	0,73	1,00	0,41	0,31	0,54	0,55	0,46	0,52	0,40	0,58
Művkif	0,34	0,41	1,00	0,31	0,59	0,32	0,48	0,56	0,52	0,55
Skicc	0,32	0,31	0,31	1,00	0,38	0,29	0,36	0,35	0,31	0,44
Választ.	0,46	0,54	0,59	0,38	1,00	0,38	0,47	0,54	0,51	0,61
Kül	0,55	0,55	0,32	0,29	0,38	1,00	0,41	0,38	0,34	0,50
Expr	0,46	0,46	0,48	0,36	0,47	0,41	1,00	0,62	0,58	0,59
Anyag	0,50	0,52	0,56	0,35	0,54	0,38	0,62	1,00	0,64	0,57
Techn	0,40	0,40	0,52	0,31	0,51	0,34	0,58	0,64	1,00	0,59
Konc	0,51	0,58	0,55	0,44	0,61	0,50	0,59	0,57	0,59	1,00

Mivel ezek a korrelációk lényegesek, a reziduálisoknak valamilyen közös faktort kell tartalmazniuk. Bár több különböző kölcsönhatás erősíti a mérési hibát a reziduálisoknál, az a sejtésünk, hogy a zsűritagok és a kollektciók közötti kölcsönhatások a dominánsak. Ezt a hipotézist alátámasztja a következő gondolatmenet is. Az adatokat az első zsűri-

kísérlet során egy úgynevezett *vertikális pontozási módszerrel* gyűjtöttük, ahol a zsűritagok egyesével az összes kritérium szerint lepontozták a kollektciókat. (Az ún. *horizontális adatgyűjtés* során a zsűritagok előbb az 1. kritérium szerint pontozzák le az összes kollektciót, majd ezután jön a 2. kritérium stb.) A horizontális pontozás összefüggést hozhat létre a különböző kritériumok szerinti pontszámok között: az első benyomás egy olyan hangulatot kelthet a kollektcióval kapcsolatban (ez tetszik nekem, ezt viszont nem szeretem), amely minden pontszámot befolyásol. Statisztikai fogalmaink szerint ez a zsűritag és a kollektció közötti kölcsönhatás. Az 5. részben elemezzük a második zsűrizés során gyűjtött adatokat, amelyeket már az említett horizontális módon végzett értékeléssel kaptunk. Az ott kapott eredmények alátámasztják megállapításainkat a vertikális adatgyűjtés hátrányairól.

6. táblázat. A zsűritag-kollektció kölcsönhatást is tartalmazó lineáris modell reziduálisainak korrelációi

	Felism	Terv	Művkif	Skicc	Választ	Kül	Expr	Anyag	Techn	Konc
Felism	1,00	0,37	-0,30	-0,20	-0,19	0,02	-0,06	-0,07	-0,26	-0,23
Terv	0,37	1,00	-0,22	-0,19	-0,07	0,04	-0,19	-0,14	-0,34	-0,14
Művkif	-0,30	-0,22	1,00	-0,04	0,10	-0,35	-0,12	-0,05	-0,01	-0,02
Skicc	-0,20	-0,19	-0,04	1,00	0,11	-0,18	-0,17	-0,08	-0,22	-0,14
Választ	-0,19	-0,07	0,10	0,11	1,00	-0,28	-0,20	-0,02	-0,12	-0,06
Kül	0,02	0,04	-0,35	-0,18	-0,28	1,00	-0,21	-0,28	-0,22	-0,14
Expr	-0,06	-0,19	-0,12	-0,17	-0,20	-0,21	1,00	0,12	0,07	-0,02
Anyag	-0,07	-0,14	-0,05	-0,08	-0,02	-0,28	0,12	1,00	0,12	-0,19
Techn	-0,26	-0,34	-0,01	0,22	-0,12	-0,22	0,07	0,12	1,00	0,04
Konc	-0,23	-0,14	-0,02	-0,14	-0,06	-0,14	-0,02	-0,19	0,04	1,00

Ahhoz, hogy becsülhessük az említett kölcsönhatásokat, úgy módosítottuk modellünket, hogy a zsűritag-kollektció kölcsönhatást is beépítettük. Ezen modellnél is kiszámoltuk a korrelációkat, mely értékeket a 6. táblázat mutatja. Látható az 5. és 6. táblázat adatainak összehasonlításából, hogy az 5. táblázatban szereplő korrelációk döntő részben a kollektció-zsűritag kölcsönhatással magyarázhatóak.

A bonyolultabb modellből adódó reziduálisok korrelációi alacsonyabb abszolút értékek és negatív értékeket is felvesznek. Amennyiben ezek a korrelációk szignifikánsan különböznek a 0-tól, annyiban ez olyan hatásoknak tudható be, amelyek nem szerepelnek a modellünkben, ilyen pl. a kritériumok és a kollektciók közötti, illetve a hármas kölcsönhatás. Ezeket azonban nem becsüljük: számunkra elegendő annak megmutatása, hogy a kombinált kritériumok alacsony megbízhatósága a zsűritagok és a kollektciók közötti kölcsönhatással magyarázható.

Összefoglalásul megállapíthatjuk, hogy az első zsűrizésnél a különböző kritériumok pontszámai nem csupán a kollektción illetve a zsűritagon külön-külön múlnak. A következményeket már elemeztük: a kritériumok és a zsűrorok közötti kölcsönhatás azt ered-

ményezi, hogy a reziduálisok korreláltak lesznek, és mivel a korreláció pozitív, a kombinált kritérium megbízhatósága (lényegesen) alacsonyabb, mint amit a Spearman-Brown formulától várnánk. A zsűrizés, mint vizsgaforma szempontjából ez azt jelenti, hogy *a kollektiókat kritériumonként kell elbíráltatni ahhoz, hogy megbízható osztályzatokhoz juthassunk*. Ez az értékelési módszer nem különösebben bonyolult és időigénye sem nagyobb, mint a hagyományos, egy kollektiót valamennyi szempont szerint egymás után értékelő módszeré. Lényege az, hogy az értékelők egy-egy kritérium szerint valamennyi munkát áttekintik és pontozzák, majd a következő kritériumra térnek rá és ismét megnézik az összes munkát ennek teljesítését vizsgálva. Ez a módszer azonnal szembeütővé teszi, ha egy tanuló jól tervez, alapos háttér-kutatásokat végez, hiszen ezeket a produktumokat is összehasonlítják a többiekével, nemcsak a főművek „mérkőznek egymással”. Az érettségén, ahol – reményeink szerint – nemcsak a művészeti vagy tervezői pályára készülők vesznek részt, a horizontális értékelésnek nagy szerepe lesz a vizuális képességek teljes spektrumának feltárásában.

A zsűritagok „szigorúsága” (homogenitás-vizsgálat a kategória súlyok lineáris struktúrájával)

A vizsgálat bemutatása

A homogenitás-vizsgálatban számszerű értéket keresünk a zsűritag és a választható kategória minden párosításához. Ezért, ha összesen 15 zsűritag pontoz és 4 kategória közül választanak, akkor összesen 60 paramétert kell becsülnünk (eltekintve a kollektiók pontszámaitól). Mivel nincs nagyon sok adatunk, ezért a becslések statisztikai stabilitása meglehetősen gyenge lesz, és ez azzal a következménnyel jár, hogy a becslések nagyrészt tetszőlegessé és nehezen interpretálhatókká válnak. Ebből a kényelmetlen helyzetből egy lehetséges kiút a paraméterekre vonatkozó feltételek kikötése. Az első megkötés az volt, hogy azt követeltük meg, hogy az egyes kategóriák kvantifikációja minden zsűritagra azonos legyen. Ezt a választást a tanulmányunkban már bemutattuk (ezek voltak az úgynevezett korlátozottan optimális pontszámok). Ennek a megközelítésnek a hátránya, hogy a megkötések túl erősek lehetnek. Ezért tapasztaltuk a jelen tanulmányban, hogy nagyon hasonló eredményeket kaptunk a korlátozottan optimális és a megfigyelt pontszámokra. Ez a rész egy másik lehetőséget vázol fel: lehetséges, hogy kevésbé szigorú megkötésekkel éljünk, olyanokkal, amelyek megmagyarázzák, hogy miért is térnek el zsűritagról zsűritagra az optimális kvantifikációk; és amelyek egyúttal a jövőbeli alkalmazások megbízhatóságát is növelik.

Az $x_i - y_l$ különbségek összegét (x_i jelöli az i . kollektió értékét, y_l pedig a zsűritagok együttese által hozzárendelt pontszámot) úgy is tekinthetjük, mint azt a mérési hibát, amely akkor lép fel, ha egy zsűritag az i . kollektiót az l . kategóriába sorolja. Természetesen abban bízunk, hogy a mérési hibák egymástól függetlenek, de elképzelhető, hogy ez mégsem teljesül.

Képzelnék el, hogy a j . zsűritag torzít abban az értelemben, hogy szisztematikusan alacsonyabb pontokat ad, mint amit a kollektciók megérdemelnek – ekkor az ettől a zsűritagtól eredő mérési hibák nem függetlenek. Ennek az oka, hogy a j . zsűritag túlságosan szigorú (az átlagos zsűritaghoz képest). Egy hasonló hatás lép fel akkor, amikor a mérési hiba abszolút értékét vizsgáljuk: vannak zsűritagok, akik túlértékelnek: kitűnőt adnak, ha a kollektció nem túl rossz, és nagyon rossz pontot adnak, ha a kollektció nem túl jó – míg mások éppen ellenkezőleg pontoznak: nagyon kis szórással adják a pontokat – minden, ami nem túl rossz 6-ot kap, és minden ami nem túl jó 5-t. Ezt a két tendenciát a mérési hiba $x_i - a_j z_i - b_j$ alakjában történő megadásával lehet reprezentálni, ahol b_j a j . zsűritaghoz tartozó szigorúsági paraméter, x_i az értékelt mű átlagos értéke, $a_j z_i + b_j$ pedig az x_i egy közelítése. Tegyük fel, hogy $a_j = 1$ és $b_j > 0$. Ez azt jelenti, hogy ha a j . zsűritag az i . kollektciót az l kategóriába helyezte, akkor a z_i átlagos számszerűsítés alulbecsli az i . kollektció értékét, és ezt az alulbecslést kompenzálja a b_j pozitív bónusz. Így a pozitív b_j paraméterek az átlagnál szigorúbb zsűritagokhoz tartoznak. Hasonló módon belátató, hogy egy nagy a_j érték ahhoz a zsűritaghoz tartozik, amelyiknek a pontszámai kicsi variabilitást mutatnak (összehasonlítva egy átlagos zsűritaggal). A – ritkán előforduló – negatív a_j érték olyan zsűritagra mutat, akinek a véleménye nem egyezik meg a zsűritagok többségével, alacsony értékű kollektciókra ad magas pontszámot és fordítva.

7. táblázat. Skála és szigorúsági paraméterek az 1. és 2. kritériumra (design feladat)

	Felismerhetőség (Felism)		Tervezőmunka (Terv)	
	a_j	b_j	a_j	b_j
15	0,34	0,33	0,30	0,38
16	1,01	-0,68	1,40	-1,15
17	0,76	-0,01	0,91	-0,13
18	0,82	0,16	0,76	0,09
19	0,72	0,02	0,78	0,10
20	0,77	0,41	0,63	0,44
21	0,46	-0,10	0,71	0,03
22	0,28	0,27	-0,17	0,18
23	1,05	-0,08	0,62	-0,06
24	0,74	0,08	0,55	-0,12
25	0,82	0,37	0,67	0,24
26	0,87	-0,16	0,48	-0,01
27	1,17	-0,59	0,85	-0,18
28	0,69	0,05	0,81	-0,09
30	0,70	0,46	0,74	0,43
31	0,39	-0,17	0,69	0,03

Ez a modell több mint 50%-kal csökkenti a reziduális szórást a legtöbb esetben. Ez megnyugtató: a modell nemcsak elegánsan megmagyarázható paramétereket tartalmaz,

hanem meglehetősen hatékony a szórás megmagyarázásában is. Ha a projekt-rendszerű, zsűrizéssel értékelt vizsga lesz az érettségi vizsgamódszer, minden bizonnyal lehetőségünk lesz a zsűrorok és a házi bírálatot végző tanárok továbbképzésére. Egy ilyen továbbképzésen lefolytatott, az itt leírthoz hasonló vizsgálattal és ezzel a számítási módszerrel visszajelezhetjük, mennyire szigorúak illetve „jólelkűek” az egyes bírálók, s ezzel bírálati módszereiket is befolyásolhatjuk. A 7. táblázatban két példa található a skála és szigorúsági paraméter becsléseire.

Ebből a táblázatból jól látható, hogy a zsűritagok jelentősen különböznek a szigorúság tekintetében. A 16. zsűritagnak van a legalacsonyabb szigorúsági paramétere, ez azt mutatja, hogy ő általában magasabb pontokat ad, mint a többiek. Ugyanakkor a skálaparaméternek ennél a zsűritagnál a legmagasabb az értéke, amely arra utal, hogy ez a zsűritag kis változatossággal használta a kategóriákat, és így szét kellett húzni a pontjait a kategóriák számszerűsítésénél ahhoz, hogy megfeleljen az átlagos pontszámoknak. Egy másik érdekes jelenség a 22. zsűritaghoz a 2. kritériumnál (tervezés) rendelt negatív skálaparaméter. Ez azt jelenti, hogy ez a zsűritag általában nem ért egyet a többiek értékelésével. Mivel a skálaparaméter abszolút értéke kicsi, így az adódik, hogy az ezen zsűritag ítéletéből adódó $a_j z_i + b_j$ pontszámok kis szóródást mutatnak.

A negatív skálaparaméter-érték a zsűritagok feladatának nehézségét mutatja. Ebből a szempontból figyelemre méltó, hogy a *design* témánál csak egyetlen negatív skálaparaméter-értéket találtunk, míg az expresszív, „*Pillantás a végtelenbe*” témánál hetet, közülük négyet a 8. számú zsűritagnál, akit így joggal nevezhetünk kiugró értéknek, vagy a legkevésbé megbízható zsűritagnak. Ez az eredmény ismét megerősíti azt a véleményünket, hogy a tervezési, *design* jellegű feladatok egyértelműbben elbírálhatóak, tehát a vizsgára alkalmasabbak, mint az expresszív feladatok.

Az x_{jin} pontszámok megbízhatósága

Az x_{jin} pontszámok megbízhatóságát a többi pontszámokhoz hasonló módon becsülhetjük. A 8. táblázat a 10 kritériumra és a kritériumok szerint vett átlagpontszámra tartalmazza a megbízhatóság becslését a *az expresszív* témánál.

A 10-ből 7 kritériumnál a megbízhatóság magasabb a *design* feladatnál, míg a három másiknál (*témafelismerés (felism)*, *skiccek (skicc)* és a *témaspecifikus (kül)* kritériumnál) itt az alacsonyabb a megbízhatóság – ezen belül a *témaspecifikus* kritériumnál jelentősen alacsonyabb. Ezen különbség megmagyarázása túlmutat a jelen elemzés lehetőségein, feltehetően arról van szó, hogy mást mér ez a kritérium mint a többi. Az *összbenyomással* és az *első benyomással* kapcsolatosan az alábbi következtetések vonhatók le.

- Az *összbenyomás* legalább olyan megbízható mint a kritériumok átlagán alapuló értékelés, sőt a *design* feladat esetén még jobb is.
- A *design* feladat értékelése megbízhatóbb mind az *összbenyomás* mind az *első benyomás* esetén.
- Az *összbenyomás* megbízhatóbb az *első benyomásnál*.
- A *design* feladat témaspecifikus kritériuma rossz. (Az adott motívumvilágból az amatőr tervezők – de minden bizonnyal a profik sem – képesek az egyes helyiség funkcióját egyértelműen jelző tapétamintát készíteni.)

8. táblázat. A megfigyelt (megf), a korlátozottan optimális (x^*) és az optimális (x) pontszámok megbízhatósága az expresszív témánál (két zsűritag esetén)

Kritérium	megf.	x^*	x_{lin}	x
Felism	0,43	0,44	0,61	0,68
Terv	0,39	0,40	0,52	0,59
Művkif	0,48	0,48	0,58	0,68
Skicc	0,39	0,44	0,61	0,65
Választ	0,35	0,36	0,49	0,52
Kül	0,48	0,53	0,64	0,73
Expr	0,42	0,42	0,55	0,62
Anyag	0,37	0,37	0,50	0,55
Techn	0,54	0,57	0,66	0,74
Konc	0,45	0,50	0,63	0,67
Átlagpontszám	0,56	0,54	0,71	

9. táblázat. A kritériumok, az átlagpontszám, a globális értékelés és az első benyomás reliabilitása a két témánál az x_{lin} pontszámok alapján

Kritérium	expresszív	design
Felism	0,61	0,56
Terv	0,52	0,53
Művkif	0,58	0,59
Skicc	0,61	0,53
Választ.	0,48	0,64
Kül	0,64	0,40
Expr	0,55	0,60
Anyag	0,50	0,63
Techn	0,66	0,69
Konc	0,63	0,69
Átlagpontszám	0,70	0,70
Összbenyomás	0,69	0,76
Első benyomás	0,64	0,70

Ezek a következtetések azonnal leolvashatók a táblázatból, de a különbségek nem nagyon meggyőzőek. A különbség azzal is mérhető, hogy hány zsűritagra van szükség egy minimális megbízhatóság eléréséhez. Ha az *első benyomás* három zsűritag véleményén alapulna, akkor a megbízhatóság 0,73 illetve 0,78 lenne az expresszív, illetve a *design* témánál. Ha a vizsgálatokat a megbízhatóságra korlátozzuk, akkor azt lehet állítani, hogy magasabb megbízhatóság érhető el, ha három zsűritagot kérünk meg arra, hogy pontozza

az *első benyomását* mintha két zsűritagot kérünk arra, hogy értékelje a kritériumokat és ezután explicit vagy implicit módon (*összbenyomás* értékelésével) átlagoljuk a kritériumonkénti pontszámokat.

Mivel az iskolai vizsgahelyzetben nincs mód 3 fős zsűri foglalkoztatására, mindent meg kell tenni annak érdekében, hogy a két fős (a szaktanárból és egy külső szakértőből álló) vizsgabizottság megbízható véleményt alkothasson. Illusztrált kritériumrendszerrel és az imént ismertetett „horizontális” értékelő módszer alkalmazásával a vizsga megbízhatósága jelentősen javítható.

A zsűritagok felcserélhetősége

A vizsga megbízhatósága szempontjából alapvetően fontos, hogy a zsűrorok felcserélhetőek legyenek. Esetünkben ez újabb kutatandó problémát vet fel: *vajon elvárható-e, hogy azonos képzettségű, hasonló szakmai gyakorlatú bírálók közel azonos módon fogjanak ítélni?* Amikor két zsűritag értékeli kollektívákat, ritkán egyeznek meg száz százalékosan. Egy fontos oka a teljes egyezés hiányának a mérési hiba, hiszen a zsűritag egy „tökéletlen mérőeszköz”, a pontszáma zajjal terhelt. A kérdés az, hogy ettől a „zajtól” eltekintve az egyik zsűritag kicserélhető-e a másikkal. Ha ez lenne a helyzet, akkor bármely zsűritagot megkérhetnénk a pontozásra, és egyenlő értékű mérési eredményhez jutnánk. Ebben az esetben azt mondjuk, hogy a zsűritagok felcserélhetőek. A zsűritagok felcserélhetősége nagyon erős követelmény. Ha a zsűritagokat egy teszt itemjeinek tekintjük, akkor ez a követelmény annak felel meg, hogy az itemek egymással párhuzamosak. Ha ez lenne a helyzet (nemcsak a mintabeli 15 zsűritagra, hanem a teljes populációra) a világ nagyon egyszerű lenne: az adódna, hogy a jövőbeli értékelések pontossága csak a zsűritagok számán múlik, nem pedig azon, hogy mely zsűritagokat használjuk. Ha viszont a zsűritagok nem felcserélhetőek, akkor ugyanannak a munkának az értékelése nemcsak a mérési hiba, hanem a zsűritagok közötti szisztematikus eltérések miatt is eltérhet (egy zsűritag enyhébben pontoz a másikkal – hasonlóan ahhoz, hogy az egyik item nehezebb a másikkal). Ahogy ezt már a 2. részben láttuk, vannak szisztematikus eltérések a zsűritagok között, hiszen különben az optimális súlyok egyenlők (vagy nagyon hasonlóak) lennének. Az a tény, hogy a zsűritagok különböznek, nagyon különböző dolgokat jelenthet. Elképzelhető, hogy a 15-ből 14 zsűritag nagyon hasonlít egymásra, viszont a 15. teljesen különbözően viselkedik. Vagy az is előfordulhat, hogy bármely két zsűritag nagyjából hasonló mértékben tér el egymástól. Ha ez utóbbi a helyzet, nincs nagy probléma. Viszont az előbbi esetben, a kiugró zsűritag (és az a része a populációnak, amelyet ő képvisel) gondot okozhat, különösen ha a jövőbeli értékelésnél csak két vagy három zsűritag fogja a diákok munkáit értékelni.

A homogenitás-vizsgálat alapján kapott optimális pontszám azt adja meg, amit az átlagos zsűritag pontszámának nevezhetnénk. Az optimális pontszámok és az egyes zsűritagokhoz tartozó pontszámok egyezésének mértékét a kollektívák összpontszámának és ezen zsűritag által adott pontszámok közötti korrelációval mérhetjük oly módon, hogy az optimális vagy a korlátozottan optimális súlyokat használjuk. Ha a zsűritagokat egy teszt itemjeiként tekintjük, akkor ez a korreláció megfelel az item-teszt korrelációnak. Minél magasabb a korreláció, az adott zsűritag annál jobban emlékeztet az átlagos zsűritagra.

Ezek a – korlátozottan optimális pontszámok és kategória súlyok alapján számolt – korrelációk találhatóak a 10. táblázatban.

10. táblázat. Item-teszt korreláció a 10 kritériumra és a 15 zsűritagra (R1–R14,R29)

	Felism	Terv	Művkif	Skicc	Választ	Kül	Expr	Anyag	Techn	Konc
R1	0,83	0,61	0,78	0,77	0,74	0,72	0,80	0,86	0,88	0,84
R2	0,76	0,66	0,66	0,59	0,76	0,92	0,73	0,80	0,64	0,66
R3	0,77	0,76	0,81	0,37	0,55	0,91	0,79	0,34	0,67	0,62
R4	0,60	0,67	0,60	0,63	0,51	0,77	0,47	0,63	0,92	0,88
R5	0,90	0,80	0,83	0,88	0,80	0,87	0,79	0,80	0,75	0,62
R6	0,70	0,54	0,83	0,80	0,67	0,89	0,71	0,62	0,57	0,55
R7	0,49	0,35	0,69	0,49	0,47	0,48	0,68	0,54	0,75	0,43
R8	0,49	0,51	0,58	0,58	0,21	0,50	0,44	0,45	0,77	0,89
R9	0,79	0,65	0,66	0,89	0,52	0,65	0,65	0,46	0,84	0,87
R10	0,72	0,39	0,60	0,91	0,22	0,78	0,63	0,54	0,67	0,73
R11	0,73	0,85	0,91	0,76	0,80	0,89	0,70	0,62	0,82	0,88
R12	0,63	0,73	0,87	0,58	0,68	0,91	0,76	0,89	0,83	0,74
R13	0,73	0,77	0,84	0,82	0,77	0,91	0,83	0,90	0,85	0,85
R14	0,94	0,86	0,81	0,74	0,65	0,61	0,60	0,50	0,81	0,69
R29	0,43	0,66	0,28	0,70	0,35	0,60	0,21	0,12	0,79	0,54

11. táblázat. A 10. táblázat korrelációinak (oszloponkénti) rangsora és a rangok átlaga

	Felism	Terv	Művkif	Skicc	Választ	Kül	Expr	Anyag	Techn	Konc	Átlag
R1	3	11	8	6	5	10	2	3	2	6	5,6
R2	6	8	11	11	4	1	6	4	14	10	7,5
R3	5	5	6	15	9	4	3	14	12	11	8,4
R4	12	7	13	10	11	9	13	6	1	3	8,5
R5	2	3	5	3	2	7	4	5	10	12	5,3
R6	10	12	4	5	7	5	7	7	15	13	8,5
R7	13	15	9	14	12	15	9	9	11	15	12,2
R8	14	13	14	12	15	14	14	13	9	1	11,9
R9	4	10	10	2	10	11	10	12	4	4	7,7
R10	9	14	12	1	14	8	11	10	13	8	10,0
R11	8	2	1	7	1	6	8	8	6	2	4,9
R12	11	6	2	13	6	3	5	2	5	7	6,0
R13	7	4	3	4	3	2	1	1	3	5	3,3
R14	1	1	7	8	8	12	12	11	7	9	7,6
R29	15	9	15	9	13	13	15	15	8	14	12,6

A 11. táblázatban a korrelációk (oszloponkénti) rangsora szerepel. Az 1 rang a legmagasabb, a 15 pedig a legalacsonyabb korrelációt jelenti. A táblázat utolsó sora az adott zsűritag átlagos rangszáma. A 10. és 11. táblázatból látható, hogy a 13. zsűritag hasonlít legjobban az átlagos zsűritagra, míg a 29. zsűritagnak van a legmagasabb átlagos rangja, és van az átlagos zsűritaggal jó néhány nagyon alacsony korrelációs értéke is. Ezen két táblázat alapján természetes a 29. zsűritagot egy lehetséges kiugró értéknek tekinteni. A 29. zsűritag (lehetséges) kiugrónak minősítése nem jelenti a 29. zsűritag személyének szakmai megítélését. Csak annyit mond, hogy a 29. zsűritag különbözik leginkább az átlagos zsűritagtól, azaz ő a legkevésbé tipikus értékelő, és ez egy semleges értékelésként tekintendő. Ez az eredmény minden bizonnyal azzal magyarázható, hogy míg a többi zsűritag legalább 10 esztendeje a pályán lévő, gyakorló pedagógus, a 29. számú zsűritag frissen végzett középiskolai rajztanár. Ő az egyetlen a mintában, akinek pedagógus diplomája mellett nem képzőművész, hanem iparművész végzettsége van. Mindebből arra következtethetünk, hogy amennyiben országos vizsgaként használjuk a zsűrizéses módszert, a központi – az iskolákba másod-értékelőként kiküldött – vizsgáztatókat képzettség és szakmai gyakorlat szempontjából körültekintően kell megválogatnunk. Fontos, hogy több éves oktatási gyakorlattal rendelkező szakemberek legyenek. Ugyanakkor ez az eredmény felveti annak szükségességét is, hogy a képzőművész – habitusú tanárokat alaposan felkészítsük a design jellegű feladatok elbírálására.

Projekt munkák „horizontális” zsűrije

Eredményeink bebizonyították, hogy a megbízhatóság további javítása csak úgy képzelhető el, hogy több kritériumról gyűjtünk adatokat, de úgy, hogy kevésbé függjenek össze. Véleményünk szerint erre a legkézenfekvőbb lehetőség az, ha a zsűritagok kritériumonként, nem pedig kollekciónként értékelnek. Ezt az elképzelést sikerült 1996. tavaszán meg is valósítani. A megelőző évben összegyűjtött kollekciókat ezúttal kritériumonként értékelte 12 illetve 14 zsűror. A legfontosabb kérdés az volt, hogy az összpontszám megbízhatósága valóban növekszik-e ennél a kísérleti elrendezésnél. A 12. és 13. táblázatban az 1995-ös és 1996-os értékelés megbízhatósági adatait hasonlítjuk össze (arra a feltételezett – reális – esetre, amikor minden kollekciót 2 zsűror értékelt).

12. táblázat. Megbízhatóság az expresszív feladatnál

Kritérium	1996	1995
A választott téma felismerhetősége (Felism)	0,57	0,43
A tervezőmunka, a szándékok felismerhetősége (Terv)	0,49	0,39
A választott kifejezőmód originalitása (Művkif)	0,54	0,48
Tervek és változatok (Skicc)	0,67	0,39
A tanuló választása a lehetőségek között (Választ)	0,43	0,35
Különleges hatás (Kül): a perspektíva alkalmazása	0,70	0,48

12. táblázat folytatása

Kritérium	1996	1995
A többi vizuális kifejezőeszköz kapcsolata az alkotói szándékkal (Expr)	0,48	0,42
Az anyag / technika megválasztása (Anyag)	0,46	0,37
Technikai tudás, ábrázoló képesség (Techn)	0,44	0,54
A téma-orientáltság, a tanuló koncentrációja (Konc)	0,65	0,45
Összpontszám	0,71	0,56

13. táblázat. Megbízhatóság a design feladatnál

Kritérium	1996	1995
A választott téma felismerhetősége (Felism)	0,69	0,62
A tervezőmunka, a szándékok felismerhetősége (Terv)	0,59	0,50
A választott kifejezőmód originalitása (Művkif)	0,74	0,48
Tervek és változatok (Skicc)	0,58	0,58
A tanuló választása a lehetőségek között (Választ)	0,61	0,46
Különleges hatás (Kül): a szobák funkciójának megf. motívum	0,60	0,45
A többi vizuális kifejezőeszköz kapcsolata az alkotói szándékkal (Expr)	0,72	0,55
Az anyag / technika megválasztása (Anyag)	0,73	0,51
Technikai tudás, ábrázoló képesség (Techn)	0,74	0,47
A téma-orientáltság, a tanuló koncentrációja (Konc)	0,75	0,59
Összpontszám	0,84	0,65

Mindkét esetben nőtt az összpontszám megbízhatósága – feltehetően annak eredményeként, hogy a kísérleti elrendezés a várakozásainknak megfelelően csökkentette a zsűror-kollekció kölcsönhatást. Érdekes különbség látszik viszont a két téma között. A design feladatnál minden kritérium és az összpontszám is jóval megbízhatóbb, ami egyértelműen abból adódik, hogy az expresszív téma megítélésénél sokkal több a szubjektív elem.

Érdeemes összehasonlítani a két értékelés során kialakult kollekció-rangsorokat (14. táblázat).

A 14. táblázatból megállapítható, hogy a zsűrorok a design feladatnál minimális eltéréssel ismét ugyanazt a rangsort állították fel, tehát egy hasonló jellegű témánál – néhány határesetről eltekintve – nagy biztonsággal eldönthető a zsűrorok pontszámai alapján, hogy mely kollekció felel meg egy adott szinten. Ez az eredmény is azt igazolja, hogy a zsűrizés megbízható értékelési módszer.

14. táblázat. A kollektciók rangsora a két értékelésnél

1995. évi hely (1996.évi hely)	Kollektció sorszáma	Összpontszám 1995 (1996) max=46
1. (4)	19	41,90 (40,1)
2.(12)	21	41,70 (35,8)
3. (2)	13	41,20 (43,5)
4. (1)	33	41,20 (43,8)
5.(13)	27	39,00 (34,8)
6. (9)	46	38,80 (37,5)
...		
52. (42)	2	19,40 (22,3)
53. (55)	40	18,60 (13,6)
54. (53)	37	18,30 (15,4)
55. (54)	42	16,20 (14,3)
56. (57)	43	14,60 (10,5)
57. (56)	48	14,10 (11,6)

Következtetések

A *projekt-jellegű feladat bírálata*, vagyis egy feladathoz készült több alkotás – vázlatok, variációk, feljegyzések, háttér-kutatások képes dokumentumai – egyidejű értékelése a magyar rajztanítás értékelési gyakorlatától alapvetően különböző, *különlegesen nehéz zsűrizési helyzet*. Eredményeink azonban igazolják, hogy ez a holland, angol és amerikai vizsgarendszerekben egyaránt használatos módszer valóban megbízható vizsgálóvá alakítható. Megállapításainkat az alábbiakban összegezzük:

- Értékelési kritérium-rendszerünk lényegében megfelelt feladatának, a kritériumok legtöbbje alkalmasnak bizonyult a projekt-feladatok értékelésére. A *kritériumok konkrétabbá tételével*, például egy értékelési jelentés kérésével, ahol a zsűritagoknak meg kell indokolniuk az adott kategória választását és *képes értékelési utasítás* adásával, vélhetően tovább javítható az értékelés megbízhatósága.
- Az *expresszív feladatoknál megbízhatóbban értékelhetők a design feladatok*. Design feladatot alkalmaz a rajz tantárgyban nyújtott teljesítmény értékelésére – a Bevezetőben bemutatott holland vizsgarendszer mellett – az angol GCSE vizsga is. Erre természetesen csak akkor van lehetőség, ha a tananyagban kellő súllyal szerepelnek a kézművesség, iparművészet, népművészet és ipari formatervezés ismeretei és a tervezési, modellezési, tárgykészítési és építési technikák. Mivel a Nemzeti Alaptanterv „Művészetek” műveltségterületének a „Vizuális kultúra” oktatásáról szóló részében a „Környezetkultúra” a három kiemelt témakör egyike (a másik kettő a Képzőművészet és a Vizuális kommunikáció – ez utóbbi szintén tartalmaz de-

sign elemeket), joggal remélhető, hogy a magyar rajztanításban is jelentőségéhez mérten segítheti a korszerű műveltségi anyag elterjedését.

- A *zsűritagok szigorúságáról* ill. megbízhatóságáról az itt közölt számításokkal, bemért (a két zsűri-kísérletben értékelt) 57 kollekciónk illetve hasonlóan bevizsgálható újabb művek segítségével a vizsgáztatókat felkészítő továbbképzéseken meggyőződhetünk az esetleges képzési hiányosságokat orvosolhatjuk és a művészet-szemlélet különbségeiben rejlő „elfogultságot” némileg csökkenthetjük.
- A zsűritagok hajlamosak a kollekciókat egyetlen itemként kezelni, ahol a kollekció globális minősége (az összbemérés) alapvetően fontos tényező a különböző analitikus kritériumokra való pontozáskor. Mivel a vizsga lényege éppen az, hogy a tanulók vizuális képességeinek széles köréről szerezzünk ismereteket, egy olyan módszert kell alkalmaznunk, amely némileg közömbösíti a kollekció összhátását.
- Az 5. részben vizsgált *horizontális értékelés* lényegében beváltotta a hozzá fűzött reményeket. Két zsűri-kísérletünk eredményeinek összehasonlítása igazolja, hogy ha a vizsgán a kollekciókat „horizontálisan”, tehát azonos kritérium szerint végigpontozva, a kollekciókat egymáshoz mérve értékeljük, lényegesen megbízhatóbb adatokhoz jutunk. Abból adódóan, hogy a zsűritagok kritériumoként, nem pedig kollekciónként értékelték, jelentősen csökkent a zsűror-kollekció kölcsönhatás és ezzel párhuzamosan nőtt az értékelés megbízhatósága – különösen a design témánál. Egyúttal kiderült, hogy a kritériumokénti értékelés eredményessége függ attól is, hogy az értékelendő kollekciók témája mekkora teret enged a szubjektív benyomásoknak.

A központilag végzett értékelés és az iskolákban adott pontszámok közötti szignifikáns eltérés mutatja a *zsűritagok központi felkészítésének* fontosságát és az analitikus megközelítés értelmét. Lehetségesnek látszik ugyanakkor az analitikus szempontok számának csökkentése és így az értékelési folyamat egyszerűsítése. Számításaink alapján a Függelékben közölt értékelési szempontrendszer könnyen optimalizálható.

A zsűrizéses vizsgamódszer gyakorlati alkalmazásához megfontolandó lehet a *szigorúságbeli és skála-paraméterbeli különbségek kísérleti úton történő háttérbe szorítása*. Két vagy három zsűrizett kollekció referenciapontként történő bemutatása segíthet az egyes zsűritagok saját skálájának felülírásában és így egy közös skála előállításában. Természetesen kérdéses, hogy ilyen előírt skálák működnek-e különböző feladatokra: tudja-e a zsűritag használni a pillantás témájú referencia-munkát a design kollekció értékeléséhez? Ez egy izgalmas kérdés, és a zsűrizett alkotások, melyeket a Magyar Iparművészeti Főiskola Tanárképző Tanszékének Vizuális Nevelési Gyűjteményében őrzünk, meglehetősen stabil referencia-pontként használhatók. A zsűrizés, mint vizsgamódszer megbízhatóságának további javítása csak úgy képzelhető el, hogy több kritériumról gyűjtünk adatokat, de úgy, hogy kevésbé fűggenek össze annál, mint amit a jelen elemzés első részében vizsgált vertikális adatgyűjtésnél tapasztaltunk.

A projekt, mint vizsgáztatási módszer egyértelműen jól szerepelt mindkét kísérletünkben. Lényegesen több információt adott a felvételi és érettségi vizsgákon jelenleg használt csendélet- vagy alakrajznál a tanulók teljesítményének olyan fontos komponenseiről, mint az originalitás, a tervezőképesség. Ez a műfaj *sokkal jobban közelíti a kortárs kép-*

zóművészet irányzatait, mint az akadémikus tanulmányrajzok. A zsűrizést követő vitán készült magnófelvételek és az írásos vélemények szerint valamennyi tanár *sokkal jobb, megbízhatóbb értékelési módnak tartja, mint a hagyományos, 3 óra alatt készült egyetlen alkotásra alapozott vizsgáztatást.* A művészeti főiskolák és a rajztanárképzés véleménye ebben a kérdésben megegyezett, ezért 1996 tavaszán a tantárgy történetében első ízben hirdethettünk eredményt OKTV szintű rajzi versenyen.³ A résztvevők projekt-módszerrel dolgoztak, vázlataikat, munkanaplójukat is mellékeltek kész művükhöz. A 12 tagú zsűri által kiválasztott legjobbak zárthelyi rajzi feladatot és művészettörténet tesztet is megoldottak, a győztesek pedig felvételi kedvezményben részesültek. Ez a verseny volt kísérleti munkánk első, sikeres gyakorlati felhasználása.

Irodalom

Kárpáti Andrea (1995): Projekt rendszerű vizsga a vizuális nevelésben. *Új Pedagógiai Szemle*, 11. sz. 18–28.

Kárpáti Andrea és Schönau, D. W. (1995): *Standardizált vizsga – Vizuális nevelés.* Országos Közoktatási Intézet, Értékelési Központ, Budapest.

Kárpáti Andrea (1996): Die Projekt-Portfolio-Methode – ein Modell der ungarischen Kunsterziehung für Prüfung und Abitur. In: Seiter, J., Reiterer, E., (szerk.): *Lust auf Kunst? Tendenzen und Perspektiven der Kunstpädagogik.* Schulheft: Wien, 95–109.

Kárpáti Andrea és Schönau, D. W. (megjelenés alatt): Towards An European Curriculum And Assessment In Art: Adaptation of The Dutch Final Examination in Art in Hungary. Közlésre elfogadva a *Journal of Multicultural and Cross- Cultural Research in Art Education* c. folyóirathoz.

Schönau, D. W. (1994): Final examinations in the visual arts in the Netherlands. *Art Education*, March 34–39.

Schönau, D. W. és Kárpáti Andrea (1996): *Standardized Examinations – Art.* CITO: Arnhem.

³ Az Országos Vizuális Verseny „Valaki, aki fontos nekem” címmel a Művelődési és Közoktatási Minisztérium és a Magyar Iparművészeti Főiskola Tanárképző Tanszéke szervezésében, Zombori Béla, a tanszék szakmetodikus munkatársa gondozásával, mintegy 400 III. és IV. osztályos gimnazista részvételével 1995 szeptemberétől 1996 márciusáig zajlott. A tanulók képzőművészet, (rajz és szobrászat), alkalmazott grafika, tárgytervezés és fotó témakörben adhatták be projektjeiket. Az országos tanulmányi versenyt már kiírták a jövő tanévre is, a téma ezúttal: „Útközben”.

ABSTRACT

ANDREA KÁRPÁTI, ANDRÁS ZEMPLÉNI, NORMAN D. VERHELST,
NIELS H. VELDUIJZEN AND DIEDERIK W. SCHÖNAU:
EXPERT AGREEMENT IN JUDGING ART PROJECTS

Between 1993–1995, a research project for the modernisation of the Hungarian system of final examination was carried out by the Hungarian Institute for Public Education in cooperation with Cito, the Dutch National Institute for Educational Measurement and the Hungarian Academy of Crafts and Design. The subjects involved were mathematics, English language, mother tongue, biology and visual arts. Beyond the adaptation of the Dutch examination methods, the aim of the authors of this paper was to improve and further develop the core element of the Dutch art examination, the jurying procedure of art and design projects prepared in the course of 3 months of the examination period. Jurying is a widely applied method of assessment in the visual arts, but we could not find any published research results on the validity and reliability of this procedure. As the new model of the final examinations in the arts in Hungary envisages the replacement of the current exam based on a set of academic drawing exercises with arts projects, it seems necessary to describe how many jurors, what sort of criteria and what form of assessing the works is necessary for reliable results. After comparative studies of educational documents and on-site observations of art teaching and examination practices by experts in both countries, we executed two experimental final examinations in art and design in Hungary. In the first trial, in 1994, 201 students from 21 schools representing 18 cities and towns in all geographical regions of the country did both the practical and the theoretical part of the Dutch final examination of that year. The 12 project tasks to choose from were unaltered – they included 6 two-dimensional tasks in graphic arts, painting, design and photography and 6 three-dimensional, plastic and design tasks. Portfolios of two projects (an expressive and a design task) were judged by a group of 37 trained jurors in the first jurying experiment and by 31 newly selected and trained jurors in the second. The analysis of the results shows that *project-based tasks*, especially design projects, can be reliably assessed if at least two, optimally three jurors are employed who base their judgements on a set of assessment criteria and not on their global impressions only. „*Vertical assessment*” – judging a work according to all criterion at the same time, the traditional method used by all fine arts competitions and exhibition jurying procedures – proved to be significantly less reliable than judging all the works according to one criterion at a time. This research proves that results of art, education can be reliably assessed through sophisticated tasks representing all genres of the visual arts not just drawing. But it also proves that traditional jurying practices need to be changed for the examination procedure. The employment of trained jurors, preferably *one external and one or two internal* (school-based) *evaluators* who compare all *project portfolios* (evidence of background studies, sketches, plans, variations and a final version of the project task) according to one given criterion at a time is necessary for reliable assessment.

Magyar Pedagógia, 97. Number 3–4. 203–234. (1997)

Levelezési cím / Address for correspondence: Kárpáti Andrea, Eötvös Loránd Tudományegyetem Neveléstudományi Tanszék, H-1146 Budapest, Ajtósi-Dürer sor 19–21.

Függelék

ÉRTÉKELÉSI KRITÉRIUMOK A HOLLAND ÉRETTSÉGI MUNKÁK ZSŰRIJÉHEZ

1. A választott téma felismerhetősége (FELISM)

- 1 A téma teljesen felismerhetetlen
- 2 A téma felismerhető, ha alaposabban szemügyre vesszük a kiállítási anyagot, de nem túl meggyőző
- 3 A téma felismerhető, ha alaposabban szemügyre vesszük a kiállítási anyagot, a választott megoldás meggyőző
- 4 A téma azonnal felismerhető

2. A tervezőmunka, a diák szándékainak felismerhetősége (TERV)

- 1 A tanuló szándékai (kiindulópontja, megközelítésmódja) felismerhetetlen
- 2 A tanuló szándékai felismerhetők, ha alaposabban szemügyre vesszük a kiállítási anyagot, de nem túl meggyőzően
- 3 A tanuló szándékai felismerhetők, ha alaposabban szemügyre vesszük a kiállítási anyagot, meggyőzően
- 4 A tanuló szándékai azonnal felismerhetők

3. A választott művészi kifejezésmód originalitása (MŰVKIF)

- 1 A választott kifejezésmód szemiatikus, a legkönnyebb utat választotta
- 2 A választott kifejezésmód néhány újdonságot tartalmaz
- 3 A választott kifejezésmód új, de nem túlzottan eredeti
- 4 A választott kifejezésmód eredeti, elmélyült munkára utal

4. A tanuló egy sor különféle megoldással kísérletezett, amelyekről skiccek és változatok tanúskodnak (SKICC)

- 0 A tanuló nem készített változatokat, variációkat
- 1 A tanuló különféle megoldásokkal kísérletezett, de munkájában kevés a változatosság
- 2 A tanuló különféle megoldásokkal kísérletezett, munkája meglehetősen változatos
- 3 A tanuló különféle megoldásokkal kísérletezett, munkája rendkívül változatos

5. A tanuló választása a megoldási lehetőségek /tervek /változatok között (VÁLASZT)

- 1 A tanuló minden kritika nélkül, találmra/megalapozatlanul választott a megoldási lehetőségek között
- 2 A tanuló meglehetősen megalapozottan választott, bár nem volt teljesen tisztában képességeivel / a lehetőségekkel
- 3 A tanuló megalapozottan, képességeinek és lehetőségeinek ismeretében választott

6. Különleges képi hatások, amelyek a feladatokhoz kapcsolódnak (KÜL)

A különleges hatások: a *perspektíva az A/ feladathoz* (Egy pillantás a végtelenbe) és a *szobák funkciójának megfelelő motívumok és színvilág kiválasztása a B/ feladathoz* (Készítsen három különféle design-mintát egy szállodába...)

6A – HASZNÁLD EZEKET A KRITÉRIUMOKAT AZ A/ FELADATHOZ !

- 0 A tanuló nem használt perspektívát
- 1 A perspektíva, mint ábrázolási konvenció nem járul hozzá a tanuló alkotói szándékainak megvalósításához
- 2 A perspektíva, mint ábrázolási konvenció kevésbé járul hozzá a tanuló alkotói szándékainak megvalósításához
- 3 A perspektíva, mint ábrázolási konvenció közepes mértékben járul hozzá a tanuló alkotói szándékainak megvalósításához
- 4 A perspektíva igen fontos eszköze a tanuló alkotói szándékai megvalósításának

6B – HASZNÁLD EZEKET A KRITÉRIUMOKAT A B/ FELADATHOZ !

- 0 A tanuló semmiféle kapcsolatot nem teremt az egyes szobák funkciója és a design mintája / színvilága között
- 1 Az egyik tapéta-terv kapcsolatban van a szoba funkciójával
- 2 Két tapéta-terv kapcsolatban van a szoba funkciójával
- 3 Mindhárom tapéta-terv kapcsolatban van a szoba funkciójával, de nem túl változatosak
- 4 Mindhárom tapéta-terv kapcsolatban van a szoba funkciójával, és a tervek különböznek egymástól, változatosak

7. A többi vizuális kifejezőeszköz, amelyet a tanuló kiválasztott (szín, kompozíció, ritmus stb.) kapcsolata a tanuló alkotói szándékával (EXPR)

- 1 A többi vizuális kifejező eszköz nem járul hozzá lényegesen a tanuló alkotói szándékának megvalósításához
- 2 A többi vizuális kifejező eszköz alig járul hozzá a tanuló alkotói szándékának megvalósításához
- 3 A többi vizuális kifejező eszköz hozzájárul a tanuló alkotói szándékának megvalósításához, a tartalom/funkció és a választott forma összhangja megfelelő
- 4 A többi vizuális kifejezőeszköz nagy mértékben hozzájárult a tanuló alkotói szándékainak megvalósításához, a tartalom/ funkció és forma tökéletes összhangban van

8. Az anyag / technika megválasztásának sikeressége a tanuló alkotói szándékainak megvalósítása szempontjából (ANYAG)

- 1 Az anyag/ technika nem kapcsolódik a tanuló alkotói szándékaihoz
- 2 Az anyag/ technika kevésbé kapcsolódik a tanuló alkotói szándékaihoz
- 3 Az anyag/ technika megválasztása hozzájárult a választott téma sikeres megvalósításához, a tartalom/funkció és a választott anyag/technika összhangja megfelelő
- 4 Az anyag/ technika megválasztása nagy mértékben hozzájárult a választott téma sikeres megvalósításához, a tartalom/funkció és a választott anyag/technika összhangja tökéletes

9. Technikai tudás, ábrázoló képesség (TECHN)

- 0 Az anyag/technika használata nem mutatja a képzés hatását, ügyetlen
- 1 Az anyag/technika használata nem kielégítő, nem illeszkedik jól a tanuló alkotói szándékaihoz
- 2 Az anyag/technika használata kielégítő, de nem illeszkedik jól a tanuló alkotói szándékaihoz
- 3 Az anyag/technika használata jó, a tanuló alkotói szándékainak kifejezéséhez hozzájárul
- 4 Igen magas színvonalú technikai megoldások, fejlett ábrázolóképeség

10. A téma-orientáltság szintje, a tanuló koncentrációja (KONC)

- 0 A kiállítási anyag tanúsága szerint a tanuló koncentrátság nélkül, belső fegyelem és irányultság nélkül dolgozott
- 1 A kiállítási anyag némi koncentrátságról tanúskodik
- 2 A kiállítási anyag megmutatja, hogy a tanuló megpróbálta megoldani a feladatot, de nem tudta azonos színvonalon végigvinni a munkát, figyelme elkalandozott, a koncentrátság színvonala változó
- 3 A kiállítási anyag erős koncentrátságról tanúskodik
- 4 A kiállítási anyag kiemelkedő koncentrátságról, összeszedettségéről, a téma iránti odaadásról tanúskodik.