The HETDEX Survey Emission-line Exploration and Source Classification[*]

Dustin Davis[1] , Karl Gebhardt[1] , Erin Mentuch Cooper[1,2] , Robin Ciardullo[3,4] , Maximilian Fabricius[5,6] ,
Daniel J. Farrow[5,6] , John J. Feldmeier[7] , Steven L. Finkelstein[1] , Eric Gawiser[8] , Caryl Gronwall[3,4] , Gary J. Hill[2,1] ,
Ulrich Hopp[5,6] , Lindsay R. House[1] , Donghui Jeong[3,4] , Wolfram Kollatschny[9] , Eiichiro Komatsu[10,11] ,
Martin Landriau[12] , Chenxu Liu[1] , Shun Saito[11,13] , Sarah Tuttle[14] , Isak G. B. Wold[15,16,17] , Gregory R. Zeimann[18] , and
Yechi Zhang[19,20]

[1] Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA
[2] McDonald Observatory, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA
[3] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[4] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[5] Max-Planck Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany
[6] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany
[7] Department of Physics, Astronomy, Geology, and Environmental Sciences, Youngstown State University Youngstown, OH 44555, USA
[8] Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[9] Institut für Astrophysik, Universität Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany
[10] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany
[11] Kavli Institute for the Physics and Mathematics of the Universe (WPI), Todai Institutes for Advanced Study, The University of Tokyo, Kashiwanoha, Kashiwa,
Chiba 277-8583, Japan
[12] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[13] Institute for Multi-messenger Astrophysics and Cosmology, Department of Physics, Missouri University of Science and Technology, 1315 N Pine St, Rolla, MO
65409, USA
[14] Department of Astronomy, University of Washington, 3910 15th Ave NE, Room C319, Seattle, WA 98195-0002, USA
[15] Astrophysics Science Division, Goddard Space Flight Center, Greenbelt, MD 20771, USA
[16] Department of Physics, The Catholic University of America, Washington, DC 20064, USA
[17] Center for Research and Exploration in Space Science and Technology, NASA/GSFC, Greenbelt, MD 20771, USA
[18] Hobby-Eberly Telescope, University of Texas, Austin, Austin, TX 78712, USA
[19] Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan
[20] Department of Astronomy, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

Abstract

The Hobby–Eberly Telescope Dark Energy Experiment (HETDEX) is an untargeted spectroscopic survey that aims to measure the expansion rate of the universe at $z \sim 2.4$ to 1% precision for both $H(z)$ and $D_A(z)$. HETDEX is in the process of mapping in excess of one million Lyα emitting (LAE) galaxies and a similar number of lower-z galaxies as a tracer of the large-scale structure. The success of the measurement is predicated on the post-observation separation of galaxies with Lyα emission from the lower-z interloping galaxies, primarily [O II], with low contamination and high recovery rates. The Emission Line eXplorer (ELiXer) is the principal classification tool for HETDEX, providing a tunable balance between contamination and completeness as dictated by science needs. By combining multiple selection criteria, ELiXer improves upon the 20 Å rest-frame equivalent width cut commonly used to distinguish LAEs from lower-z [O II] emitting galaxies. Despite a spectral resolving power, $R \sim 800$, that cannot resolve the [O II] doublet, we demonstrate the ability to distinguish LAEs from foreground galaxies with 98.1% accuracy. We estimate a contamination rate of Lyα by [O II] of 1.2% and a Lyα recovery rate of 99.1% using the default ELiXer configuration. These rates meet the HETDEX science requirements.

1. Introduction

It is generally acknowledged that the universe is expanding and that the expansion is accelerating. Although surprising at the time, the accelerated expansion has come to be the consensus understanding since the early work of Perlmutter et al. (1999), Riess et al. (1998). Since then, many observations have confirmed and refined the measures of this expansion with such increased precision that a possible tension may have emerged in the results from the various broad measurement camps (Aloni et al. 2022; Di Valentino et al. 2021, among others). Regardless, whether this tension is a consequence of real physics, as yet unidentified systematics, or some combination, we are essentially limited to only two anchor points, one from the recent past (~ 72 km s^{-1} Mpc^{-1}; Riess et al. 2009, 2022; Dhawan et al. 2018; Mörtsell et al. 2022, and others) and one from the epoch of recombination (~ 67 km s^{-1} Mpc^{-1}; Alam et al. 2017; Aiola et al. 2020; Planck Collaboration et al. 2020a, and others), from which to constrain descriptions of dark energy. Further understanding

---

requires additional data points from different epochs in the expansion history of the universe. Multiple efforts are in progress to provide those data, including the following, far from exhaustive, list: the Dark Energy Survey (DES; The Dark Energy Survey Collaboration et al. 2016), the Baryon Oscillation Spectroscopic Survey (Dawson et al. 2013), the extended Baryon Oscillation Spectroscopic Survey (Alam et al. 2021), the Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009), Euclid (Laureijs et al. 2011), the DESI Survey (DESI Collaboration et al. 2016; Dey et al. 2019), and, of course, the Hobby–Eberly Telescope Dark Energy Experiment (HETDEX; Ramsey et al. 1998; Gebhardt et al. 2021; Hill et al. 2021).

HETDEX is a multiyear untargeted spectroscopic survey designed to make new measurements of the Hubble parameter, $H(z)$, and the angular diameter distance, $D_A(z)$, at $z \sim 2.4$ to better than 1% accuracy in an effort to better characterize dark energy and look for possible evolution. HETDEX observations fall into two large, high galactic latitude fields. The $\sim$390 deg^2 "spring" field is centered near (R.A., decl.) 13^h00^m $+53^d00^m$, and the $\sim$150 deg^2 "fall" field is centered near 1^h30^m $+0^d00^m$ (Gebhardt et al. 2021). Functionally, HETDEX seeks to map the 3D positions of some 10^6 galaxies between $1.88 < z < 3.52$ and use their large-scale clustering to derive $H(z)$ and $D_A(z)$. More specifically, the galaxies HETDEX is using for large-scale structure are identified by their bright, conveniently redshifted into the optical, Lyα emission lines. These Lyα emitters (LAEs) are generally small, blue, rapidly star-forming galaxies that, while uncommon in the local universe, are present in large numbers in the HETDEX redshift search window (Partridge & Peebles 1967; Gawiser et al. 2007; Nilsson 2007; Finkelstein 2010, and many others).

The HETDEX Visible Integral-Field Replicable Unit Spectrographs (VIRUS; Hill et al. 2021) cover the wavelength range 3500–5500 Å with $R \sim 750$–900, and are optimized to detect Lyα flux down to $\sim 4 \times 10^{-17}$ erg s^{-1} cm^{-2} (increasing to closer to 2×10^{-16} erg s^{-1} cm^{-2} at the extreme blue end of the range). This allows the detection of Lyα luminosities down to about $10^{42.3}$ erg s^{-1} for $z \sim 2.4$. Since it is of utmost importance to know the redshift of the observed galaxies, the emission must be correctly identified. However, the relatively narrow wavelength range often limits our ability to capture multiple emission lines, and the low spectral resolving power prohibits most doublet splitting, making classifications difficult. Around 95% of HETDEX emission line detections[21] are spectra containing only one, apparently single peaked (given the HETDEX spectral resolving power) emission line, and Lyα is not the only emission line to fall into this observed wavelength range. Neutral hydrogen (and dust) in each source galaxy's Interstellar Medium and in the Intergalactic Medium along our line of sight effectively eliminates emission lines blueward of Lyα at higher redshifts (Haardt & Madau 1996; Cowie & Hu 1998; Meiksin 2006; Overzier et al. 2012; Vanzella et al. 2018), leaving low-z galaxies as the primary contaminate to be considered.

In the relatively nearby universe, intrinsically small, line-emitting faint galaxies can be misidentified as their higher-

redshift cousins. In particular, at the low HETDEX spectral resolving power and with no strong lines in the wavelengths around it, the [O II] 3727 Å emission line can be confused with Lyα 1216 Å, which similarly appears unique in its spectral neighborhood. In a common case, HETDEX observations detect only a single, fairly narrow emission line and little or no continuum at the detection limits. Most likely the line is either Lyα and originates from a high-z galaxy, or [O II] from a low-redshift interloper, and unfortunately, these two primary cases occur in roughly equivalent numbers (Adams et al. 2011; Gebhardt et al. 2021). Since the HETDEX $H(z)$ and $D_A(z)$ measurements are sensitive to interloper clustering (Leung et al. 2017; Farrow et al. 2021; Gebhardt et al. 2021), the contamination from [O II] in the LAE sample needs to be $\lesssim$2% (Gebhardt et al. 2021). Historically, a 20 Å equivalent width cut (using the rest frame of Lyα) has been used to segregate [O II] from Lyα (Gronwall et al. 2007; Adams et al. 2011), and indeed, this criterion is quite effective. However, used by itself, the discriminant can still lead to >4% contamination and degrade the recovery of lower equivalent width Lyα lines (Acquaviva et al. 2014). Leung et al. (2017) improves on the 20 Å cut by taking a Bayesian approach and including information on the luminosity functions and equivalent width distributions of Lyα and [O II]. From their modeled data, they report an expected contamination by [O II] of between $\sim$0.5% and 3.0% at a cost of $\sim$6.0% to 2.4% lost LAEs, depending on the methods used. This is a significant enhancement over the simpler 20 Å cut, and in this work, we are able to extend and improve on Leung et al. (2017) by (1) incorporating additional selection criteria, (2) considering other emission lines as contaminants, and (3) comparing directly against observational data.

The HETDEX Emission Line eXplorer (ELiXer) software incorporates and extends these classification works, integrates supplemental data and additional classification criteria, and expands the analysis to consider more than two dozen other emission lines. Its primary objective is to classify every HETDEX emission line detection by assigning the correct redshift to the observed emission lines. In addition to its primary function as an emission line classifier, ELiXer also provides diagnostic and data integrity checking to supplement that of the HETDEX pipeline (Gebhardt et al. 2021), which is run prior to the ELiXer invocation and provides the detection coordinates, observation conditions, the calibrated and point-spread function (PSF) weighted spectra, emission line parameter measurements (flux, line width), and CCD information as ELiXer inputs. These features are useful for identifying and debugging some issues (e.g., errant sky subtraction, stuck and/or hot pixels, amplifier interference, etc.) as well as in the manual inspection of individual detections.

While ELiXer does classify all HETDEX detections regardless of magnitude, additional classification support is provided for continuum-bright sources via another software tool utilized by HETDEX called *Diagnose*, developed for the Hobby Eberly Telescope VIRUS Parallel Survey (G. Zeimann et al. 2023, in preparation). For a further description of source classification and redshift assignment of HETDEX sources, please see Mentuch Cooper (2023). Here, however, we focus only the bulk of the HETDEX detections, where ELiXer is the primary (or only) classifier. For this work, we reference ELiXer version 1.16 used in the generation of the most recent

---

[21] HDR3 is limited to emission line detections with signal-to-noise ratio (S/N) $\geqslant$ 4.8, of which 95% have only a single detected emission line. The fraction of detections with only a single line is partly a function of the S/N cut and other selection criteria used to define a sample. As in Mentuch Cooper (2023), S/N $\geqslant$ 5.5 is commonly used as it is effectively free from noise detections (Section 5.4). For S/N $\geqslant$ 5.5, 70% of HETDEX spectra consists of only a single emission line, and the entire sample is reduce by 60%.

HETDEX detections catalog, HETDEX Data Release 3 (HDR3). This catalog contains more than 1.5 million entries and was released internally in 2022 April with a public version to be released in the future. We report a projected HETDEX LAE contamination rate from [O II] of 1.2% ($\pm$0.1%) and an additional 0.8% ($\pm$0.1%) from all other sources, along with an LAE recovery rate of 99.1% ($\pm$3.3%) for the default classification configuration. ELiXer provides a tunable Lyα classifier, allowing the balancing of contamination versus completeness as needed for specific science goals (see Section 4.4). ELiXer is a work in progress and continues to evolve and improve as more data are collected, both from HETDEX and from other surveys, and as classification methods are added and refined.

The remainder of this paper is organized as follows: Section 2 provides an overview of the various photometric catalogs currently included in ELiXer. Section 3 describes the classification methodologies and supporting functions. Section 4 covers the selection of a spectrocopic-z assessment sample (SzAS) providing spectroscopic redshifts from various imaging catalogs and the results of testing against that sample. Section 5 presents a discussion of the results and the science implications. Section 6 summarizes the work and future enhancements. Example ELiXer detection reports are shown in Appendix with descriptions provided for the major features.

Throughout the paper, the Planck 2018 cosmology (Planck Collaboration et al. 2020b) with $\Omega_\Lambda = 0.69$, $\Omega_m = 0.31$, and $H_0 = 67.7$ km s^{-1} Mpc^{-1} is assumed. All magnitudes are in the AB system (Oke & Gunn 1983), and the coordinates are J2000.

2. Imaging Catalogs

HETDEX is an untargeted spectroscopic survey, and the spectra alone provide most of the critical information for object classification. Coupled with the on-sky positions of the associated fibers, these data form the basis for the HETDEX cosmology measurements. For the brighter detections, a source's redshift and, to a lesser degree, its physical extent and morphology can be determined securely from the spectra. However, for the fainter emission line detections, additional information from archival photometric imaging, including an object's magnitude, color, angular and/or physical size, morphology, and even on-sky neighbors, can prove quite useful in ascertaining its identity. Even superimposing the HETDEX fiber positions on imaging data can provide diagnostic checks on the astrometry and the reduction pipeline. Given these substantial benefits, ELiXer attempts to match all HETDEX observations with multiband archival photometry at the highest angular resolution and imaging depth available.

2.1. Individual Catalog Summaries

At the time of writing, ELiXer references 11 separate imaging catalogs, most with their own associated object catalog. These catalogs are of varying depth, resolution, band-coverage, and footprint. Additional catalogs can be added at any time and, several new or expanded source lists are anticipated before the next HETDEX data release. With the exceptions of an r-band survey, HSC-DEX, from the Hyper Suprime-Cam and a g-band survey, KPNO;HETDEX-IM, from Kitt Peak National Observatory (KPNO) that were specially designed and executed for HETDEX, all imaging

and object catalogs are archival and publicly available. These catalogs are summarized in Table 1 and in the list below.

1. *Canada–France–Hawaii Telescope Legacy Survey (CFHTLS).* A multiband (*ugriz*) imaging survey and joint venture of the National Research Council of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii utilizes the MegaPrime/MegaCam on the 3.6 m Canada–France–Hawaii Telescope (CFHT) on Maunakea. ELiXer uses the deep and wide fields, D3/W3 centered near R.A. 210°, decl. +52° (Cuillandre et al. 2012).

2. *Hubble Space Telescope (HST) Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) in the Extended Groth Strip (EGS).* CANDELS is a deep HST survey (900+ orbits) with multiple filters in the optical (using the Advanced Camera for Surveys, hereafter ACS) and near-IR (using the Wide Field Camera 3, hereafter WFC3) studying on galaxy evolution with an emphasis on cosmic dawn and cosmic high noon. The EGS is one of the five fields of CANDELS and is centered near R.A. 215°, decl. +53° (Grogin et al. 2011; Koekemoer et al. 2011; Stefanon et al. 2017).

3. *HST Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) in the Great Observatories Origins Deep Survey, North (GOODS-N).* Another of the 5 CANDELS fields (see previous bullet), GOODS-N is centered near R.A. 189°, decl. +62° (Dickinson et al. 2003; Grogin et al. 2011; Koekemoer et al. 2011; Barro et al. 2019).

4. *Hyper Suprime-Cam HETDEX Survey (HSC-DEX).* This survey consists of three nights of HSC r-band observations with the Subaru/HSC in 2015–2018 (PI: Andreas Schulze) and 2019–2020 (PI: Shiro Mukae) and covers the $\sim$250 deg^2 area of the HETDEX spring field. Data reduction and source detections were performed with version 6.7 of the HSC pipeline, hscPipe (Bosch et al. 2018), and produced r-band images with a 10σ limit of $r = 25.1$ mag in a 2″ diameter circular aperture. These HSC r-band images are complementary to the existing imaging data of the Kitt Peak 4 m Mosaic camera and the CFHT Wide-Field Legacy survey.

5. *Kitt Peak National Observatory HETDEX Imaging Survey (KPNO; HETDEX-IM).* A g-band survey with the Mosaic camera on the Mayall 4 m telescope at KPNO in 2011–2014 (PI: Robin Ciardullo).

6. *Cosmic Evolution Survey (COSMOS) with Dark Energy Camera (DECam).* The 3 deg^2 *ugriz*-band COSMOS DECam catalog was generated with the same procedure used for the larger field-of-view Spitzer/HETDEX Exploratory Large-Area DECam survey listed below (Wold et al. 2019). This also overlaps with Laigle et al. (2016).

7. *Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP).* This is a multidepth, multiband, wide-field imaging survey using the Hyper Suprime-Cam on the 8.2 m Subaru at the Maunakea Observatories. For HDR3, ELiXer uses HSC-SSP Public Data Release 3 from 2021 August (Aihara et al. 2022).

8. *Spitzer/HETDEX Exploratory Large-Area with Dark Energy Camera (DECam).* This survey covers 17.5 deg^2 of the HETDEX fall field within the Sloan Digital

Table 1
Summary of the Imaging Surveys Incorporated into ELiXer

Name	HETDEX Field	Overlap[a]	Filters and Depth[b]	PSF FWHM[c]	Object Catalog[d]
Canada–France–Hawaii Telescope Legacy Survey (CFHTLS)	Spring	4%	Deep: $u(26.3)$, $g(26.0)$, $r(25.6)$, $i(25.4)$, $z(25.0)$ Wide: $u(25.2)$, $g(25.5)$, $r(25.0)$, $i(24.8)$, $z(23.9)$	$0\farcs6$–$1\farcs0$	phot-z
HST Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) in the Extended Groth Strip (EGS)	Spring	<1%	ACS/WFC: F606W, F814W WFC3: F105W, F125W, F140W, F160W	$0\farcs08$	spec-z, phot-z
HST Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) in the Great Observatories Origins Deep Survey, North (GOODS-N)	Spring	<1%	ACS/WFC: F435W, F606W, F775W, F814W WFC3: F105W, F125W, F160W	$0\farcs08$	spec-z, phot-z
Hyper Suprime-Cam HETDEX Survey (HSC-DEX)	Spring	44%	$r(25.5)$	$0\farcs6$–$1\farcs0$	mag only
Kitt Peak National Observatory HETDEX Imaging Survey (KPNO; HETDEX-IM)	Spring	20%	$g(24.4)$	$1\farcs1$–$1\farcs5$	mag only
Cosmic Evolution Survey (COSMOS) with Dark Energy Camera (DECam)	Fall	2%	$g(25.5)$, $r(25.5)$	$0\farcs7$–$1\farcs0$	(1) phot-z (Laigle+2015) (2) mag only
Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP)	Fall	29%	Deep $g(27.5)$, $r(27.1)$, $i(26.8)$, $z(26.3)$, $y(25.3)$ Wide $g(26.5)$, $r(26.1)$, $i(25.9)$, $z(25.1)$, $y(24.4)$	$0\farcs6$–$1\farcs0$	mag only
Spitzer/HETDEX Exploratory Large-Area with Dark Energy Camera (DECam)	Fall	25%	$u(25.4)$, $g(25.1)$, $r(24.7)$, $i(24.0)$, $z(23.7)$	$0\farcs7$–$1\farcs0$	mag only
Dark Energy Camera Legacy Survey (DECaLS)	Spring and Fall	17%	$g(24.0)$, $r(23.4)$, $z(22.5)$	$1\farcs2$	No
Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)	Spring and Fall	<1%	$g(23.3)$, $r(23.2)$, $i(23.1)$, $z(22.3)$, $y(21.3)$	$1\farcs0$–$1\farcs3$	No
Sloan Digital Sky Survey (SDSS) DR16	Spring and Fall	<1%	$u(22.0)$, $g(23.1)$, $r(22.7)$, $i(22.2)$, $z(20.7)$	$1\farcs3$	spec-z, phot-z

Notes.

[a] Fraction of HETDEX Data Release 3 within each catalog footprint, except for DECaLS, Pan-STARRS, and SDSS, which report only the fraction that does not also overlap with a previously listed catalog. Because multiple catalogs overlap, the column sums to >100%.

[b] Approximate average AB depth over the whole catalog as reported, typically for point sources and $2''$ apertures. For some g and r filters and some image tiles, ELiXer uses its own estimated depths at $1''$ and $2''$ apertures. Not all surveys use the same SDSS $ugriz$ filters, although for this purpose they are approximately similar. Only filters used by ELiXer are listed.

[c] Typically in r band.

[d] If not "no," also has an object catalog used by ELiXer with at least g or r magnitudes. Spec-z and/or phot-z redshifts are available where noted, but not necessarily for all object entries.

Sky Survey (SDSS) "Stripe 82" region. The *ugriz*-band DECam catalog is *riz*-band-selected and reaches a 5σ depth of $\sim$24.5 AB mag for point sources (Wold et al. 2019).

9. *Dark Energy Camera Legacy Survey (DECaLS)*. A multiband (*grz*) photometric survey, part of the DES (The Dark Energy Survey Collaboration et al. 2016), is based at the Cerro Tololo Inter-American Observatory using the DECam on the 4 m Blanco telescope. ELiXer uses Data Release 9, which also includes observations from the Beijing-Arizona Sky Survey and the Mayall *z*-band Legacy Survey (Dey et al. 2019).

10. *Panoramic Survey Telescope and Rapid Response System (Pan-STARRS)*. Specifically, Pan-STARRS1 (PS1) is a set of wide-field synoptic imaging surveys using the 1.8 m PS1 optical telescope at the Haleakala Observatories. PS1 collected data from 2010 through 2014 (Chambers et al. 2016).

11. *Sloan Digital Sky Survey (SDSS)*. This multiband (*ugriz*) wide-field survey in operation since 2000 uses a 2.5 m optical telescope at the Apache Point Observatory. ELiXer uses Data Release 16 from SDSS Phase-IV (Ahumada et al. 2020).

2.2. ELiXer Aperture Photometry

ELiXer directly uses the photometric imaging to gather aperture magnitudes for the HETDEX detections. While magnitudes are computed for each available filter, only *g* and *r* magnitudes are used in the classification process (Section 3). For each HETDEX detection, ELiXer identifies the catalogs with overlapping imaging and gathers postage-stamp ($9'' \times 9''$ by default) imaging cutouts centered on the HETDEX detection's coordinates. Three sets of aperture magnitudes are then computed using the Python packages *Astropy* (Astropy Collaboration et al. 2022), *Photutils* (Bradley et al. 2022), and *Source Extraction and Photometry* (SEP; Barbary 2016). The identified aperture(s) are used later to provide continuum estimates (Section 3.2) and size information (Section 3.5.1).

First, ELiXer computes a magnitude for a dynamically sized circular aperture. We center the circular aperture on the HETDEX coordinates, compute the magnitude within the aperture, and allow the aperture to grow until the magnitude stabilizes (e.g., Howell 1989). The initial size is set by a combination of the median seeing and pixel scale of the catalog +filter and is typically $\sim$1$''$ in diameter. The magnitude within the aperture is computed, with the background determined from an annulus 2$\times$ to 3$\times$ the defined maximum allowed object aperture (6$''$ diameter by default, for an annulus of 12$''$ to 18$''$). The aperture is then grown in steps of 0.$''$1, with each measurement recorded, until the maximum diameter is reached. The smallest aperture size where the magnitude change to the next step up is less than 0.01 is assigned, and the corresponding magnitude is selected.

Next, ELiXer uses SEP (Barbary 2016), which is based on the original *Source Extractor* (Bertin & Arnouts 1996), iterating over each cutout and records the magnitude, barycentric position, major and minor axes, and orientation of each identified object. ELiXer also computes and records the angular separation from each barycenter to the HETDEX coordinates and the separation to the nearest point on the bounding ellipse if the HETDEX position lies outside that ellipse. The object with the nearest barycenter to the HETDEX position whose bounding ellipse includes the HETDEX position is considered the best aperture match. If no object's ellipse includes the HETDEX position, then the object with the nearest ellipse point to the HETDEX position but no more than 0.$''$5 away is selected as the best match. If no object meets these criteria, then no SEP found object is selected, and the best circular aperture (see previous paragraph) is used for the aperture photometry.

Lastly, at each SEP identified barycenter, ELiXer computes and records the background subtracted magnitude in a fixed, 3.$''$0 diameter circular aperture. These aperture magnitudes are intended for use in any fixed-aperture spectral energy distribution-fitting and color comparisons, but are not otherwise significantly used in the core ELiXer processing.

2.3. Catalog Counterpart Matching

ELiXer also attempts to match each HETDEX detection to one or more objects in each imaging catalog with a particular focus on *g* and *r* magnitudes, which can provide additional measures for use in other ELiXer functions. Object matching is based on a combination of barycenter position and agreement between the magnitudes reported by each catalog, the magnitudes computed within the ELiXer ellipses (Section 2.2), and the HETDEX spectrum estimated *g*-band magnitude.

The nearest catalog object to the HETDEX position that falls within the selected best aperture (Section 2.2), or the nearest catalog object within 1.$''$0 of the HETDEX position if no object falls within the best aperture, is identified as the catalog match object. If the candidate object's reported magnitude is not compatible with the magnitude estimated from the HETDEX spectrum, then the next nearest object is evaluated until a match is found or the distance criteria are no longer satisfied. Compatibility with the HETDEX *g* magnitude (Section 3.2.1) is defined as an absolute difference of 0.5 mag; if the HETDEX *g* magnitude is fainter than the HETDEX magnitude limit (about 25_{AB}), then no faint-side restriction is imposed. On the other end, if both the counterpart and the HETDEX magnitudes are brighter than 22_{AB}, they are considered compatible. For the purposes of this comparison, *g* and *r* are considered equivalent. There is at most one catalog match object per catalog+filter combination. This object is later used for additional information, including spec-*z* and phot-*z* assignments if available, in the classification process.

3. Classification

Classifications in ELiXer are broadly interpreted as the identification of the redshifts of observed astrophysical objects. This properly requires the additional steps of correctly associating an observed spectrum with a single host object and furthermore identifying or bounding what constitutes that *single object*. More fundamentally, given a spectrum and a specified emission line in that spectrum, what we hereafter call the *anchor line*, ELiXer attempts to determine the identity, and thus the redshift, of that anchor line. Classification proceeds from the assumption that the anchor line is real and not spurious noise, an instrument or software artifact, or a misinterpretation of spectral data, such as the misidentification of continuum between two closely separated absorption troughs. ELiXer initially assumes that the spectrum represents a single object (single redshift); though later analysis explores

Table 2
Emission Line Candidates

Name	Rest-λ (Å)	Name	Rest-λ (Å)
O VI	1035	Hη	3835
Lyα	1216	[Ne III]	3869
N V	1241	Hζ	3889
Si II	1260	(K) Ca II[a]	3934
Si IV	1400	[Ne III]	3967
C IV	1549	(H) Ca II[a]	3968
He II	1640	Hϵ	3970
C III]	1909	Hδ	4101
C II]	2326	Hγ	4340
Mg II	2799	Hβ	4861
[Ne V]	3346	[O III]	4959
[Ne V]	3426	Na I	4980
[O II]	3727	[O III]	5007
		Na I	5153

Notes. Possible identifications for spectral lines found in the HETDEX spectra.
[a] Fit as an absorption line.

the possibility that an HETDEX spectrum is a blend of spectra from discrete but immediately adjacent or overlapping sources on the sky (within a single, common detection aperture) at different redshifts.

The focus of ELiXer's classification is placed on distinguishing Lyα from [O II], by far the most common Lyα contaminant in HETDEX data, and the bulk of the tests and conditions target that objective. Additional checks, described throughout this section, attempt to refine this bifurcated classification and identify the spectral line(s) as any one of those listen in Table 2. As will be discussed in Section 5, these *other* lines are encountered much less frequently than Lyα and [O II], and while they can be more challenging to identify, the HETDEX cosmology science is extremely robust against contamination from these misclassifications.

The classification of HETDEX detections is organized to answer three increasingly general questions, with each answer incorporating the results of the previous question. First, closely following the work of Leung et al. (2017), we evaluate the relative likelihood that the target emission line is Lyα and rather than [O II] (Adams et al. 2011; Farrow et al. 2021; Gebhardt et al. 2021). This is largely based on measurements of the emission line luminosity and equivalent width evaluated against luminosity and equivalent width distributions of Lyα and [O II] emitting galaxies from other publications interpolated at the redshift corresponding to the emission line wavelength (see Section 3.4). Second, we determine the confidence of the initial classification by performing checks against more than two dozen other emission lines. Here a weighted voting scheme is used with many independent (or semiindependent) rules applied to measured and derived features of the spectrum and detection object. Third, we assign, with some rough measure of quality, the redshift and thus the specific identity of the emission line(s). This final step incorporates some additional rules and weights to combine all prior results.

Broadly, ELiXer classifications build up evidence in a series of steps and then weighs the evidence to make a determination. The high-level steps are fairly serial and often largely independent, with their results only combined toward the end of the process. These major steps are described in more detail, and in roughly the same order, in the subsections that follow.

1. Find, fit, and score all emission and absorption lines, and set the anchor line.
2. Evaluate all combinations of found spectral lines for compatibility with redshifts, based on relative positions, strengths, etc.
3. Collect additional (aperture) photometric imaging information and any reported magnitude, spec-z, and phot-z measurements for the target object and its neighbors from non-HETDEX catalogs (Table 1).
4. Evaluate spectra shape, lines, and imaging for consistency with known astrophysical objects (star, white dwarf, active galactic nucleus, hereafter AGN, meteor, low-z galaxy).
5. Examine HETDEX data for corruption, pipeline artifacts, and instrument issues.
6. Test the compatibility of the anchor line with Lyα.
7. Perform evaluations on the anchor line, including spectral and photometric information, to specifically distinguish Lyα from [O II].
8. Perform separate evaluations on the anchor line, including spectral and photometric information, for consistency with lines other than Lyα and [O II].
9. Combine all evaluations to determine and rank likely redshifts and line classifications.
10. Reevaluate redshift classification based on clustering with ELiXer results from the other neighboring HETDEX detections.

The figures in this section illustrating some of the voting criteria and thresholds pull their data from the spectroscopic-z assessment sample (SzAS) whose selection and composition are described in Section 4.

3.1. Line Finder

Emission (and absorption) line detection is implemented as both a layered, untargeted search and a targeted line fit assuming an *anchor* line. More details will follow in the next subsections, but briefly put, the untargeted search scans the full width of the spectrum from blue to red, marks the locations of possible emission line centers, and attempts to fit a single Gaussian (in agreement with the measured instrumental resolution; Hill et al. 2021) to each position. The targeted search uses a single previously identified emission line (from the HETDEX input, user input, or the previous untargeted search) as an anchor and then assumes that anchor line is one of roughly two dozen potential emission lines (Table 2) and attempts to fit a Gaussian to the positions where other emission lines could be found, assuming that identity for the anchor line. The descriptions that follow are couched in terms of emission lines, as that is the primary use. A limited use of absorption lines is implemented and is described in Section 3.1.6.

3.1.1. Untargeted Search

The untargeted search scans the entire 1D HETDEX spectrum to identify the positions and model the parameters of potential emission lines. It is used to (1) identify the strongest line as the reference or anchor line when no initial emission line is explicitly provided, (2) mark strong lines for consistency checks with redshift solutions and to help identify blended spectra, and (3) mark line positions for follow-up visual inspection, without respect to the selected solution.

Because Markov Chain Monte Carlo (MCMC) fits are relatively computationally expensive, and HETDEX spectra typically have only one or very few emission lines, we do not want to perform such fits at each pixel along the spectrum. Instead, we first conduct a quick examination to narrow the potential locations of emission lines. We do this using two independent algorithms and then combine the output positions into a single list for further examination.

Two passes through the algorithms of this untargeted search are conducted. The first execution uses the native 2 Å binned HETDEX spectrum and focuses on identifying the common narrow spectral features. The second execution is performed after passing the original spectrum through a median filter (by default using a 5 pixel kernel), to smooth out some of the noise. This helps identify the candidate emission lines that are wider than the $\sim$400 km s^{-1} resolution of the VIRUS spectrographs and may have small noise peaks within their overall broad shape.

The first algorithm searches for the basic shape of an emission feature, a general rise to a peak and then a decline. Due to the unavoidable noise in the data, the spectra are not smooth, and the use of the first derivative to find zeros (and the second derivative to distinguish between an emission and absorption) results in more false detections than real spectral features. Instead, we look for the general shape of the lines (a rise and fall in the flux of minimum height over a minimum width), based on the spectral resolution, flux limits, and noise of HETDEX. The sets of contiguous pixels that are sufficiently wide in the spectral direction and have the expected rise-peak-fall pattern are recorded as possible emission lines, and their line centers are recorded.

The second algorithm counts contiguous pixels with flux values above some multiple of the corresponding noise (typically S/N $>$ 3, under the assumption that the flux uncertainty is distributed normally). Where the contiguous count of pixels above this noise is greater than some count (here, typically 3–5 pixels), the position of the highest flux value within that range is recorded as the possible emission line center. Essentially, this is just an S/N-cut over the spectrum. Unlike the first algorithm, the shape of the flux above the S/N-cut is irrelevant.

The line centers from each algorithm are then passed to fitting (Section 3.1.3) and scoring routines (Section 3.1.4). When model fits to the flux at those positions are successful and the computed line score is sufficiently large, the feature is recorded to a list of potential spectral lines.

After both the standard and broad-line searches are conducted, the lists of potential emission and absorption lines are merged into a single list, and any neighboring lines with line centers within in 4 Å of each other are combined into single entries by keeping only the feature with the largest line score.

As a brief note, although this is not the normal operation of ELiXer under HETDEX, if no anchor line is specified for the spectrum to be classified, the line (emission or absorption) with the largest score (Section 3.1.4) found in this untargeted search is assumed as the anchor line. If the untargeted search fails to identify any spectral lines, the wavelength bin with the largest flux value is assigned as the anchor line position.

3.1.2. Targeted Search

Unlike the untargeted search described above, the targeted search does not scan for potential emission or absorption lines, but instead attempts to fit for an emission or absorption feature at a specified position. Essentially, ELiXer attempts to fit spectral lines from a predefined list of common lines (Table 2) at their expected observed wavelength positions given an assumed identity or redshift for the anchor line. The redshift assumptions come from alternately interpreting the anchor line as each of the common lines and from any matching spectroscopic or photometric catalogs with a possible counterpart to the HETDEX detection. With each redshift assumption, all other lines in the subset that could occur within the HETDEX spectral window are fitted, allowing for some error in the systemic redshift (see *position capture* under Section 3.1.3). This is often redundant with the untargeted search in that, for higher S/N lines, the lines found in the targeted search are also found in the untargeted search. However lower S/N lines, [O III] λ4959 for example, can be missed in the initial sweep of the untargeted search. Fitting to a specific wavelength location helps avoids such misses.

Each successfully fitted line for each assumed identity of the anchor line is scored (Section 3.1.4) and associated with the redshift solution (Section 3.3) for that identification.

3.1.3. Line Fitting

ELiXer uses a simple, 4-parameter (A, μ, σ_{Line}, y) single Gaussian as the model to fit emission and absorption features:

$$F(\lambda) = \frac{A}{\sigma_{\mathrm{Line}} \sqrt{2\pi}} \exp\left(-\frac{(\lambda - \mu)^2}{2\sigma_{\mathrm{Line}}^2}\right) + y, \qquad (1)$$

where $F(\lambda)$ is the flux per 2 Å wavelength bin, A is the area under the curve or equivalently the integrated line-flux, μ is the line center, σ_{Line} is the measure of width, y is the vertical offset, or flat continuum level, and λ is the wavelength (at the midpoint of a 2 Å wide wavelength bin).

The flat continuum is a reasonable simplification, as no assumption is made as to the object type or its redshift, most HETDEX detections have continua at or below the survey's continuum flux limit, and those objects with continua bright enough to have a shape typically have multiple emission lines or are too bright to support a Lyα classification. This continuum estimate can be highly uncertain, especially for the noisier spectra, but as discussed later, multiple continuum estimates are combined to improve the uncertainty, and for the nondetections, the resulting equivalent width estimates are lower limits that favor a low contamination Lyα selection, at the cost of some completeness.

Type I AGN may have broad lines that are not well fitted by a single Gaussian (Liu et al. 2022). Such detections are marked by ELiXer with warnings, but are not confused with the fainter, compact LAEs that the software is designed to identify. We note, however, that it is possible that the simple emission line search can completely fail to find rare, extremely broad emission lines, as $\sim$3500 km s^{-1} is the maximum FWHM that ELiXer attempts to fit.

More complex models, including the fitting of multiple emission and absorption lines within a single spectral feature, have proven to be either unreliable, too computationally costly,

and/or of limited utility for the main goal of simply identifying redshifts when the vast majority of line detections are well fit by the simple, single Gaussian model. Fitting for an emission line doublet would be useful in the effort to distinguish between Lyα and [O II]; however, given the low spectral resolving power of VIRUS, $\lambda/\Delta\lambda \sim 800$ (Hill et al. 2021), the [O II] doublet (3726, 3729 Å) is unresolved as are most other doublets (Mg II (2796, 2803 Å) is sometimes marginally resolved). The increased run time of fitting these extra parameters is not justified. For smaller data sets, such as for the case of AGN exploration, more complex fitting is warranted (Liu et al. 2022), but left to those specialized projects. For ELiXer's classification needs, a description of the spectral feature that is limited to its position (wavelength), equivalent width (approximate integrated line flux and local continuum), and line width is sufficient. Additional parameters, such as the model's skewness and kurtosis, and conditions combining those and other parameters have been explored but have not been found to improve the identification of real spectral features or aid in the classification, and are thus excluded from further discussion in this work.

With the exception of the anchor line on which an MCMC fit is always performed, if a least square (LSQ) model fit passes its quality checks, no MCMC fit is conducted. This is due to the increased runtime cost of MCMC fitting weighed against the relatively modest needs for classification. In all MCMC cases however, an LSQ fit is performed first, and its results are used as initial conditions (with appropriate randomization) for the MCMC algorithm. ELiXer uses the Python *scipy* package and its *scipy.optimize.curve_fit* (Virtanen et al. 2020) as the LSQ fitter; the MCMC fitter is from the Python *emcee* package (Foreman-Mackey et al. 2013). The uncertainties in the LSQ fit are estimated using the square root of the diagonal of the covariance matrix. The uncertainties in the MCMC fit are estimated using the 68% confidence interval in the parameter distribution.

A series of loose checks evaluates the quality of each fit as minimally good, marginal, or poor. Poor fits are rejected; good fits are scored (see Section 3.1.4) in preparation for building solutions. Marginal solutions from the LSQ fitter are passed to the MCMC algorithm for improved optimization and reevaluated. If the subsequent MCMC fit is good, the fit is scored and made eligible for inclusion in redshift solutions. If the MCMC fit is not sufficiently improved over the LSQ fit, it is rejected. The quality checks include following conditions:

1. *Peak capture.* As a basic check, should the peak of the model fail to reproduce the most extreme measured data value near the line center within 50%, the fit is rejected. If the model is within 25% and 50% of the most extreme value, it is flagged for an MCMC fit. Should that MCMC fit fail to be within 25%, the fit is rejected, and no line is assumed to be at that position.

2. *Position capture.* If the fitted line center is greater than a configured maximum distance (in Å) from the local data extremum, the fit is rejected. The maximum distance allowed can depend on the assumed line identification and its assumed position, with greater separations allowed for Lyα, which can be significantly offset from the systemic redshift (Shapley et al. 2003; McLinden et al. 2011; Verhamme et al. 2018; Gurung-López et al. 2021, among others). During the untargeted search, no

variations are allowed, and a default of 8 Å ($\sim$500 km s^{-1} in the HETDEX spectral range) is used.

3. *Width capture.* If the fitted line width (here parameterized as σ) is less than 1.0 Å, i.e., significantly below the HETDEX spectral resolution of $\sim$2.0 Å (Hill et al. 2021), or if the line width is greater than the configured maximum value of 17 Å ($\sim$2700 km s^{-1} FWHM) or 25 Å ($\sim$3500 km s^{-1} FWHM) for special, broad fit attempts, the fit is rejected.

4. *Area error.* If the error on the line area (as estimated from the square root of the diagonal of the LSQ fit's covariance matrix or the 68% confidence interval on the MCMC fit) is larger than the absolute value of the area (allowing for absorption or emission), the fit is rejected.

5. *Local uniqueness.* This is used only in combination with other conditions. An emission or absorption line is considered unique if there is at most one other data extremum greater than 90% of this line's peak between $1 \times$ FWHM and $1 \times$ FWHM + 10 Å to either side of the line center. This is an alternate rough measure of local noise and is used primarily as a filter with low S/N lines.

6. *S/N and χ^2.* ELiXer uses the following definitions of S/N and χ^2:

$$S/N = \frac{\sum \sqrt{(F(\lambda) - y)^2}}{\sqrt{\sum (\text{error}^2)}}; \quad (2)$$

$$\chi^2 = \sum \left(\frac{\text{data} - \text{model}}{\text{error}}\right)^2, \quad (3)$$

where the summations are over the wavelength bins within $\pm 2\sigma$ of the fit line center. $F(\lambda)$ and y are from Equation (1). The model is the fitted flux evaluated at each corresponding wavelength bin for the data, and the error is the uncertainty on the data.

The uncertainty on the S/N is computed via standard error propagation using the MCMC or LSQ uncertainties on each of the model's Gaussian parameters.

If the LSQ fit is marginal given the previous conditions, it is rejected if (1) the S/N is less than 5.0 or (2) if the S/N is between 5.0 and 15.0, and the χ^2 is greater than 2.0. These indicate poor fits to possibly noisy data and are generally not worth pursuing. Otherwise, the S/N and χ^2 are recorded for use in line scoring.

3.1.4. Line Scoring

Every successfully fitted emission and absorption line receives a score based only on its own properties, without consideration to the position or properties of any other fitted emission or absorption lines. If that score exceeds a minimum threshold, the line, with its score, is accepted into a list of potential line candidates for later use in the redshift solution finder (Section 3.3). The minimum threshold is configurable and is set, by default, to an empirically determined value based on the manual examination of many tens of thousands of observed spectra and a simulation of spectra drawn from median HETDEX noise properties (Section 3.1.5). The redshift solutions that fit multiple lines to the spectrum receive a separate *solution score* (Section 3.3) that is based, in part, on these individual *line scores*.

The line score attempts to capture and quantify features beyond just the S/N, which is a less than ideal metric for broad

emission lines fitted with a single Gaussian. The line score takes into account additional data including the magnitude of the integrated (fitted) line flux, the line position relative to expectations, and the uniqueness of the line within a local spectral region. The intent is to codify not just the presence of each potential emission line, but the consistency and significance of that line with respect to the spectrum at an assumed redshift.

The line score calculation is defined as

$$S_L = \frac{S_{\rm lim} \cdot AN \cdot UN \cdot F_\lambda \cdot m_\sigma \cdot m_{\rm pix}}{1 + |\delta dx_0|} \qquad (4)$$

where the following statements are true:

1. S_L is the numerical line score. Noise peaks receive scores in the low single digits, typically less than 3.0. Weak emission lines (low S/N, low line flux) typically receive scores in the 5.0–15.0 range. Extremely bright, high S/N lines can even exceed a score of 100.0, but are clipped to a maximum of 100.

2. $S_{\rm lim}$ is the maximum allowed fitted S/N from a Gaussian fit, up to a configurable limit (20.0 by default). This helps scale the scoring by capping the maximum contribution of the S/N.

3. AN is the *above noise* factor, defined by the measured flux value of the emission peak divided by a noise estimate at that position and normalized by a configurable factor (by default, 5). The noise estimate used here is the standard deviation of the 3σ clipped fluxes at the same wavelength over all (448) fibers on the detector. The value of AN is clipped to the range [0, 3].

4. UN is an estimate of how unique the line is relative to the nearby spectrum (i.e., the presence of several similarly narrow, low flux peaks in the same wavelength range likely indicate noise in the spectrum). This is an encoding of the *local uniqueness* described in the previous subsection. If the candidate line is sufficiently broad, with a fit FWHM of greater than 6.5 Å or if fewer than 3 possible lines are found, the current candidate line is considered sufficiently unique, and UN takes on a value of 1; otherwise it takes on a value of 1/2.

5. F_λ is the Gaussian fitted, continuum subtracted integrated line flux in units of 10^{-17} erg s^{-1} cm^{-2}. There is no particular significance to these units; they are simply used so that the value of the line score is generally in the range of 1–100.

6. m_σ encodes the minimum acceptable Gaussian fitted σ. Values of σ greater than 1 Å result in $m_\sigma = 1$, but values less than 1 Å receive a multiplicative penalty equal to the σ value as they are unlikely to have been fit to a real emission line. This is equivalent to min $(\sigma, 1)$.

7. $m_{\rm pix}$ encodes the minimum acceptable number of pixels ($N_{\rm pix}$) over which the S/N of the line is calculated. If the number of pixels is less than $N_{\rm min}$ (by default, 10 pixels to either side of the wavelength bin containing the line center), there is a multiplicative penalty imposed equal to $N_{\rm pix}/N_{\rm min}$. Low numbers of pixels in the S/N measurement may be due to masked or invalid pixels or a line location near the edge of the wavelength range. This is equivalent to min $(N_{\rm pix}/N_{\rm min}, 1)$.

8. δdx_0 is the offset, in Å, of the fit line center from the expected location of the center line. For features found by

the untargeted search (Section 3.1.1), this is the bin with the maximum (minimum, for absorption) flux within the spectrum slice being used to fit the line. For corroborating features as part of the *targeted search* (Section 3.1.2), it is the expected position of the assumed feature for the given redshift.

An adjustment is made to the S_L if the fit S/N is less than 8.0, and the χ^2 is greater than 3.0. These are considered marginal fits that could have a large score due to the integrated line flux. In these cases, the score is reduced by a factor of $(\chi^2 - 1)$.

If the center of an emission line falls within a prominent sky-line, specifically those centered at 3545 Å or 5462 Å, and if the FWHM does not extend past the sky-line, the score is further reduced by a factor of 2, encoding the risk that the emission line is a relic of incomplete sky subtraction.

For very broad lines (fit FWHM > 20 Å), the scoring is modified by rejecting the line (setting the S_L to 0) if the fitted S/N is less than a minimum threshold (by default, 19), and the χ^2 of the Gaussian model is greater than a maximum (by default, 1.5). These fits tend to be poor, and caused by either artifacts in the data or the merging of multiple spectral features.

Since the focus is on faint galaxies with continuum below the HETDEX sensitivity, absorption features do not factor strongly in classification for most HETDEX catalog objects. As such, their base scoring value is scaled by a factor of 1/2 and optionally limited to a maximum value

3.1.5. Spectra Simulation and P(Noise)

As part of the scoring and in an effort to quantify the probability that a fitted line is simply the product of noise, we use the line-finding code to analyze simulated spectra, treating all identified emission lines as false positives (FPNs). The procedure is applied only to emission lines, not absorption lines, but the results are applicable to both.

As part of the configuration for ELiXer, we compute the PSF-weighted spectral uncertainties versus wavelength from 10^4 random, noncontinuum detections from the entire HETDEX catalog, and generate the median uncertainty for each wavelength bin. We then simulate 10^4 spectra, randomly drawing a flux for each wavelength bin (1036 random draws per spectrum over the range, 3470–5540 Å) according to the median uncertainty, and assuming a normal distribution about each uncertainty and no correlated noise between wavelength bins. Each simulated spectrum is passed through the line-finding code, and all identified emission lines are recorded with their line scores (Section 3.1.4). The line scores are binned in steps of 1.0 and normalized by the number of simulated spectra. This represents the simulated estimate of the probability that an emission line in a given scoring bin is the product of noise. This probability, P(Noise), monotonically decreases with increasing line score. Note that it is possible by this mechanism for a scoring bin to have a value of P(Noise) greater than 1.0, and that is the case for the lowest-scoring bins. For such cases, the probability is cropped to 1.0, and any emission line with a score that fall in those bins is considered to be noise. The higher-scoring bins are cropped once the P(Noise) falls below 5×10^{-4}, with that P(Noise) assumed for all emission lines with line scores above that value.

When applied to line detections in real data, any line score below the lowest score for the bin is assumed to be noise and is

rejected, and any line detection with a score above the highest score receives the $P(\text{Noise})$ of the highest score for the bin. These $P(\text{Noise})$ estimates the factor in the solution scoring (Section 3.3), described later.

Since the $P(\text{Noise})$ is based on the line scoring and on the uncertainties in the HETDEX PSF-weighted spectra, any reformulation of the line scoring or any change to the HETDEX pipeline that results in a change in flux uncertainties necessitates a recomputation of this mapping.

3.1.6. Absorption Lines

As called out by its name, ELiXer is primarily designed to identify and act on emission lines. Continuum bright HETDEX detections ($g < 22$) are also analyzed with an independent software package (*Diagnose*, G. Zeimann et al. 2023, in preparation). Nevertheless, ELiXer does currently include a limited use of absorption lines, triggered either explicitly at its invocation or automatically for the detections with continuum greater than 2×10^{-17} erg s^{-1} cm^{-2} Å^{-1}. The same untargeted search (Section 3.1.1) used for emission lines is executed for absorption lines, with the exception that the spectrum is first inverted by subtracting all the flux densities from the maximum flux density of the spectrum. This allows the fitter to treat the absorption lines as if they were emission lines, but only for purposes of line identification within the spectrum. The actual fitting (Section 3.1.3) and initial scoring (Section 3.1.4) are performed on the original, noninverted spectrum, with the appropriate sign changes to account for the different direction in the Gaussian model. And like the case for emission lines, the positions of absorption lines with scores above a configurable threshold are also marked in the 1D spectrum.

While there are 26 emission lines checked by ELiXer, only the Ca II 3968,3934 Å absorption lines are explicitly fitted and used in spectral redshift identification. Additionally, these two lines are fit simultaneously and must appear together. If they occur at the edge of the spectral range, such that only one line could be found in the spectrum, the fit is not allowed. A simple assertion is made to the pair of lines, requiring them to be of similar flux and FWHM such that the difference in flux and FWHM must be with 50% of the mean of their mean values. If the assertion fails, the fit is rejected. If the assertion passes, the lines are both accepted and contribute to the solution scoring (Section 3.3).

3.2. Continuum Estimates

Much of the classification effort rests on an accurate measure of the emission line equivalent width, so a robust estimate of the continuum underlying the emission line is of major importance. There are several, independent and semiindependent estimates of the continuum that contribute to a single combined estimate.

Since most of the independent estimates arise from photometric imaging, we calibrate our continuum-derived classification properties (described later in this section) to the bandpass continuum estimates, all of which assume a flat spectrum over the bandpass with no emission or absorption line masking (see Sections 3.2.1–3.2.4). This means we are slightly biased to overestimate the continuum level. This is more pronounced for objects such as AGN with strong, broad emission, but given the objective of accurate classification, this is a nonissue with these objects being a rare subset of HETDEX data and unlikely to be confused with the typical, continuum faint LAE. In the general case that ELiXer is designed to address, our objects have faint or undetected continuum and a single, faint emission line, so the bandpass overestimate is minimal and serves as an upper limit.

All continuum estimates from broadband photometry assume a flat spectrum point source over the bandpass and convert the magnitude to flux density at the emission line's observed wavelength rather than the filter's effective wavelength as follows:

$$f_\lambda = c \; \lambda^{-2} \times (3631 \times 10^{-23}) \times 10^{-0.4m} \qquad (5)$$

where f_λ is the flux density at the observed wavelength (in erg cm^{-2} s^{-1} Å^{-1}), c is the speed of light in the vacuum (Å s^{-1}), λ is the fitted, observed wavelength center (Å), and m is the g or r magnitude. The literal constant is in units of erg cm^{-2} s^{-1} Hz^{-1}. As most of the HETDEX emission line detections either have only r coverage or are undetected in the imaging even when multiple bands are available, a color correction to the photometric continuum estimate is rarely possible. In limited testing where photometric detections are made in both g and r, no improvement in the classification performance and no change in the classification rates are found, and so no color correction is included in this version of ELiXer.

3.2.1. HETDEX Spectrum

The HETDEX spectrum covers the entire g bandpass and therefore can be used to estimate an object's g-band magnitude without the use of external data. Sky and background subtraction is very good, and the continuum level is consistently measurable $\lesssim 10^{-18}$ erg s^{-1} cm^{-2} Å^{-1} (Gebhardt et al. 2021). We use two methods to derive the g magnitude from the HETDEX 1D spectrum. The first multiplies the HETDEX spectrum through the SDSS g filter's throughput curve using the Python *speclite* package (Kirkby 2020). ELiXer runs 1000 realizations of the HETDEX spectrum, sampling over the flux errors, and assigns the biweight (Beers et al. 1990) of those realizations to define an estimated g magnitude and its 68% confidence interval. The second method sums the total flux in the HETDEX spectrum, again with propagated errors, and uses the mean flux density and an $f_{\lambda,\text{eff}}$ of 4726 Å to set a continuum and the g-band magnitude. In both cases, the object is assumed to be a point source. The combined continuum mean is converted into a g magnitude for ease of use and comparison to other catalog-reported magnitudes.

While this estimate is reported as computed, it is used internally with an imposed flux density limit of 5.38×10^{-19} erg s^{-1} cm^{-2} Å^{-1} ($g = 25$). When our measured HETDEX continuum flux density is at least $1.2\times$ brighter than the limit, it receives the highest weight ($4\times$ standard) in the combined estimate (Section 3.2.4), as it is based on the same data that provide the line flux estimate. All other continuum estimates are from other data sources and matched by proximity. As the limit is approached, the weight rapidly drops to the standard vote weight and is considered a nondetection once the limit is reached.

A second estimate of the continuum is obtained using the y offset from the Gaussian fit to the emission line (Equation (1)). While this is the estimate nearest the emission line, it can also have a large uncertainty, and the simple Gaussian model does not allow for asymmetric line flux or different continuum levels

on either side of the line. When this estimate is brighter than the HETDEX limit, it receives a small, empirically set weight of $0.2\times$ the standard vote; otherwise it receives zero weight and is not included in the combined continuum estimate.

A third and final estimate is also recorded, but is not, by default, included in the combined continuum estimate. In this estimate, the continuum is still assumed to be flat in f_ν, but all emission and absorption lines identified in the spectrum are masked at $\pm2\sigma$ from the fitted line centers. The mean of the unmasked fluxes, with standard error propagation, is converted into a flux density and returned as the continuum estimate. With the exception of the continuum bright objects with multiple, broad spectral lines mentioned earlier, this estimate is not significantly different from the *speclite* result, and its inclusion in the combined estimate would be both redundant and somewhat inconsistent, given the other photometric estimates. It is, however, used internally in some diagnostic checks.

3.2.2. Aperture Photometry

The g- and r-bandpass continuum estimates come directly from run-time aperture photometry as described in Section 2.2. When an SEP aperture matches that of the HETDEX detection, its magnitude is used. If no SEP aperture is a match, then the smallest, stable ELiXer circular aperture provides the magnitude estimate. In either case, if the computed magnitude is fainter than the imaging limit, that limit is used, and the continuum value is flagged as a nondetected upper limit.

Since the HETDEX emission lines appear in the g band, an optional correction is allowed for translating an r-band continuum estimate to the g band; however this is not used by default, as an examination of g and r continuum estimates where both are available from the same instrument for the same objects shows no consistent trend. Additionally, Leung et al. (2017) finds no advantage in using g over r, and their simulated data actually suggest that LAE/[O II] segregation is slightly improved with r, although this is not confirmed with the observed spectra in this work.

If the measured aperture magnitude is brighter than the limiting magnitude of the image, it receives a full (1.0) weight in the final, combined estimate. If the measured aperture magnitude is fainter than the limit, it is treated as a nondetection, and the limit is used in the combined estimate. When the *limit* is used for the aperture magnitude, the weight in the combined estimate is scaled down linearly from 1.0 to 0.0 as the limit grows brighter from 26_{AB} to 24_{AB}, and a nondetection in that increasingly bright limit provides less and less useful information (noting that the HETDEX spectra has a magnitude limit near $g = 25$). The 26_{AB} and 24_{AB} boundaries selected to roughly cover the magnitude range of maximal LAE and [O II] galaxy g magnitude overlap in HETDEX.

3.2.3. Catalog Counterpart

Lastly, if a catalog counterpart can be matched to the HETDEX detection (Section 2.3), its reported bandpass magnitude (again, only g or r) is added to the list of continuum estimates. A minimum 20% flux uncertainty is assumed, even if no uncertainty is reported by the catalog. All catalog-reported values are assumed to be a proper detection and receive a full (1.0) weight.

3.2.4. Combined Continuum

The combined estimate is produced using the weighted mean of a subset of the individual continuum estimates, described in the immediately previous subsections, with less informative estimates and extreme outliers removed from consideration.

At most, a single upper limit estimate is allowed in the subset and is selected as the deepest (faintest) upper limit. This is typically the limit from the deepest photometric imaging where there is no detection or where the aperture magnitude is fainter than the image's limit. No upper limit is included if there exists a positive aperture detection. If there are three or more continuum estimates in the subset, a fairly aggressive clip is applied, which excludes the most extreme estimate(s) with values greater than $1.5\times$ the weighted biweight scale (Davis et al. 2021) while retaining a minimum subset size of two. The final combined continuum estimate is then the weighted mean of the surviving continua in the subset:

$$\bar{f}_\lambda = \frac{\sum_i (f_{\lambda_i} \ w_i \ \sigma_i^{-2})}{\sum_i w_i \ \sigma_i^{-2}}; \tag{6}$$

$$\Delta\bar{f}_\lambda = \sqrt{\frac{\sum_i (w_i \ \sigma_i^2)}{\sum_i w_i}}, \tag{7}$$

where $\bar{f}_\lambda$ is the combined (*averaged*) continuum estimate, f_{λ_i} is an individual continuum estimate, w_i is the associated weight, and σ_i is the associated standard deviation. The error, $\Delta\bar{f}_\lambda$, is the square root of the weighed average of the variances.

This defines the distribution over which the continuum is sampled for the $P(\text{LAE})/P(\text{O II})$ classifier in the next subsection.

3.3. Redshift Solutions

Distilled to its most basic functions, ELiXer's raison d'être is to assign the correct redshift to every detection as the operative analog to the classification of the target emission line. The core approach to this objective is the testing and ranking (or scoring) of many possible redshift solutions. Clearly the most secure, and consequently the highest-scoring, solutions are those with multiple identified spectral lines consistent with known rest-frame features at an assumed redshift. ELiXer's initial set of redshift solutions is generated by iterating over the lines in Table 2 and assuming, in turn, that each one represents the target emission line identification (note that the H&K absorption lines are handled differently per Section 3.1.6). With each assumed redshift, ELiXer attempts to fit all in the list, and accumulates a total solution score based on the number and quality of the successes (Section 3.1.4). At this stage, only the relative line positions are considered, with flux ratios, required lines, and other criteria considered in later steps. The more lines that are found, the more robust the solution. Unfortunately, only about 5% of ELiXer classifications are established with more than one identifiable emission line, so additional methods must be applied to confidently identify the target emission lines and assign the corresponding redshift.

3.3.1. Catalog Redshift Match

When ELiXer matches an HETDEX detection to one (or more) catalog objects (Section 2.3) that have associated spectroscopic and/or photometric redshift assignments, that information is evaluated in the context of the emission and

absorption lines identified in the HETDEX spectrum. The catalog supplied redshift, with its error, is applied to the target emission line, and all other ELiXer identified lines and the resulting rest-frame wavelengths are checked for consistency with those in Table 2. If the catalog redshift results in rest-frame wavelength matches, it boosts any previously assigned ELiXer score (Section 3.3) for that redshift, with a larger weight given to spec-z ($+100$ to the redshift solution raw score, Section 3.3.5) than to phot-z ($+5$ to the redshift solution raw score). If an ELiXer redshift solution for that catalog redshift does not exist, one is created and scored in the same way. Approximately 0.1% of the HDR3 detections has a catalog-matched spec-z counterpart, and 1.5% have a phot-z counterpart.

3.3.2. Large Galaxy Mask

In addition to matching redshift catalogs, ELiXer also compares the sky position and wavelength of each detection against an internal HETDEX catalog of large galaxies. We define this large galaxy catalog by searching the most recent versions of the Third Reference Catalog of Bright Galaxies (de Vaucouleurs et al. 1991)[22] and the UGC (Nilson 1973)[23] galaxy catalogs for objects larger than $1'$ in diameter within our survey area. In total, we find 644 large galaxies in the spring field, and 447 in the fall field. For each system, we adopt the catalog's basic parameters for position, position angle, ellipticity, and D25 semimajor axis (i.e., the size of the galaxy defined by its B-band isophote at 25.0 mag arcsec^{-2}). Prior to inclusion in the large galaxy mask, each galaxy is manually inspected to confirm that these values are reasonable. Where values of these parameters are uncertain, they are corrected to values listed in the NASA/IPAC Extragalactic Database[24] or through visual inspection of the galaxy in SDSS g-band images. Any HETDEX detection falling within $3\times$ the D25 isophotal radius of a large galaxy is tested against the spectral features expected for the system's redshift. This matching is performed in exactly the same way as for the catalog matching in the previous section, except that the scoring is scaled inversely by the distance in multiples of D25. The overall area of this large galaxy mask is dominated by a handful of nearby galaxies (NGC 5457 and NGC 4258 in the spring field, and IC 1613 and NGC 474 in the fall field).

3.3.3. Special Handling for [O III]

The [O III] 5007 Å line can be problematic to identify by equivalent-width-based methods when other oxygen or Balmer lines are not detected as it can have a large equivalent width and appear similar to Lyα. Low-z compact star-forming galaxies, planetary nebulae (PNe), extragalactic H II regions, and the outer star-forming regions of resolved galaxies could sometimes have detectable [O III] 5007 Å, but with [O III] 4959 Å, [O II] 3727 Å, and Hβ that do not reach the threshold for a standard HETDEX detection. Such objects could be classified as Lyα by the base algorithms. To protect against such misclassifications, additional tests are needed.

For observed emission lines redward of 5007 Å, but without any other nominally detected emission feature, a lower

---

[22] Available at: http://haroldcorwin.net/rc3/.

[23] https://heasarc.gsfc.nasa.gov/W3Browse/galaxy-catalog/ugc.html

[24] http://ned.ipac.caltech.edu

threshold for emission line detection is allowed at the expected positions of [O III] 4959 Å, [O II] 3727 Å, and Hβ. If one or more of those lines are detected at this reduced stringency, a redshift solution is created with a score of at least the minimum acceptable threshold, and a flag is set for follow-up manual inspection.

If one or more of the above lines are found and there is no identified imaging counterpart, a flag is also set to indicate that this could be a PN, either in the Galaxy or in intergalactic space. Given the HETDEX lines of sight are out of the plane of the Galaxy, the likelihood of encountering Galactic PNe is reduced but is certainly not zero, and several known Galactic planetaries are located in the HETDEX footprint. Given their physical proximity, most of these objects will have sizes of several arcminutes, and we test for this by looking for large spatial clusterings of emission at 5007 Å. When found, these regions are masked from use in HETDEX cosmology. A potentially more pernicious issue is PNe in the halos of nearby galaxies and intergalactic PNe within galaxies groups and clusters. These could be misinterpreted as background LAEs; though this risk is ameliorated via the check against the large galaxy mask (Section 3.3.2) and neighbor clustering (Section 3.7). Conversely, this comes at a (small) cost of the loss of some background LAEs with observed Lyα redshifted to match the [O III] 5007 Å line of on-sky adjacent foreground galaxies.

We note that [O III] 5007 Å makes up only 1% of the SzAS detections, and none are misidentified by ELiXer.

3.3.4. Object Classifications Labels

Based on combinations of spectral features (with examples given later in this subsection), some HETDEX detections are assigned classification labels. These labels indicate only that a detection is consistent with the class of object indicated by the label within the parameters defined for that class. Classifications are not mutually exclusive and are applied simply if the corresponding conditions are met. If none of the specific classification conditions are met, then no extra classification label is applied to the detection. The classification is not Boolean, but is scored, with the strength of the classification based on the number and quality of the conditions that are met. A negative classification can also be made if the failure to meet conditions is sufficiently extreme such that a classification is excluded (i.e., if the detection's properties are grossly inconsistent with the given classification).

Strongly consistent object classifications can be used to increase the score of the corresponding redshift solution, while strongly inconsistent classifications decrease the score of the corresponding solution. In this way, the object classification *can* modify the P(Lyα) result (Section 3.5) by altering the score of a multiline solution available to the P(Lyα) routines. However, the conditions are relatively strict, and the overall impact of labeling is small, with only $\sim$4% of detections actually meeting the conditions to receive an object classification label.

Additionally, a few generic labels are applied for ELiXer detections that are associated with a unique object in a photometric catalog (Section 2). These labels are only provided as suggestions and do not impact the scoring of the multiline solutions.

The ELiXer assigned labels are as follows:

1. *AGN ("agn").* The "agn" label is set if an HETDEX spectrum contains (possibly broadened) emission lines consistent with those seen in AGN. These reference emission lines are as f: O VI (1035 Å), Lyα (1216 Å), N V (1241 Å), Si II (1260 Å), Si IV (1400 Å), C IV (1549 Å), He II (1640 Å), C III] (1909 Å), C II (2326 Å), Mg II (2799 Å), and [O II] (3727 Å). For some pairs of lines, bounds on relative line fluxes must be met and certain lines must be present to support the identification of other lines. For example, if a line assumed to be C IV is observed at 5000 Å, then a line for Lyα must also be found at 3295 Å and it should be at least as strong and have a similar FWHM as C IV. If no line is observed at 3295 Å or if the feature is much weaker than the assumed C IV line, then the identification is inconsistent with that of an AGN and the C IV solution receives a reduced score.

2. *Low-z galaxy ("lzg").* The "lzg" logic is largely the same as the "agn" but with a different set of reference lines: [O II] (3727 Å), Hη (3835 Å), Hζ (3889 Å), Hε/ionC2 (3970 Å), Hδ (4101 Å), Hγ (4340 Å), Hβ (4861 Å), [O III] (4959 Å), and O III] (5007 Å). As with AGN, some bounds on line strengths must be met. For example, if a line assumed to be [O III] 5007 Å is observed at 5300 Å, then another line at 5249 Å must be observed at one-third the strength. Similarly, for HETDEX detections with strong continuum, if an absorption line is assumed to be calcium H at 3968 Å, calcium K at 3934 Å must also be present with at a similar equivalent width. If these criteria are satisfied, then the detection will be labeled "lzg." Moreover, an additional label of "o32" will be assigned to objects with an [O III] 5007 Å to O II 3727 Å flux ratio greater than 5:1.

3. *Meteor ("meteor").* With any wide-field, long-term survey, meteor intrusions on the extra-galactic observations are inevitable, and if not identified, they can be a significant nuisance source of emission (and sometimes of continuum) detections. A combination of methods are used to identify meteors in the detection catalog (Mentuch Cooper 2023).

   Since ELiXer processes only single detections in isolation, its meteor identification methodology focuses on the transient nature of the phenomenon and their fairly distinctive emission line signatures. To identify a meteor emission, we divide a spectrum into 9 nonoverlapping, noncontiguous regions by wavelength (in Å) where meteor emission lines are common: [3570, 3590], [3715, 3745], [3824, 3844], [3852, 3864], [3926, 3942], [3960, 3976], [4210, 4250], [4400, 4450], and [5160, 5220]. For the visually confirmed meteors in HETDEX, these regions often include bright features from Mg (3832, 3838, 5172, and 5183 Å) as well as typically fainter emission from Al, Ca, and Fe. Spectra that contain multiple emission lines that are within these ranges and are detected in only one of the three dithered exposures used for an observation are labeled as meteors.

4. *White dwarf ("wd").* The white dwarf label logic is very basic and simply looks for the hydrogen series absorption lines for DA and DAB types, the Helium series for DB types, and carbon and oxygen for DQ types. Additionally, to be classified as a white dwarf, the spectrum must have a blue spectral slope. Since the shape and width of the absorption features are not taken into account, nor are the presence of other features (such as pronounced H and K (Ca II) lines), it is possible to mislabel a main-sequence star, particularly an A-type, as a white dwarf. However, given the high Galactic latitude of the HETDEX survey, we do not expect the set of HETDEX detections to contain many early-type stars.

5. *Catalog labels ("gal," "star," "agn").* These are recorded as suggestions when matched to an external photometric catalog, but they do not influence any of the ELiXer logic. For example, an "agn" label from a photometric catalog matched to an HETDEX detection is considered separately from the ELiXer "agn" label logic described above and will appear in the classification labels even if the ELiXer spectral features analysis does not result in an "agn" label.

3.3.5. Redshift Solution Scoring

Each redshift solution receives three scores, a raw score, a (normalized) fractional score, and a scale score, so that the solutions can be rank ordered and assessed in terms of their viability. The raw score is the unweighted sum of the individual line scores (Section 3.1.4) of the spectral lines included in the solution, excluding the anchor line (which is common to all solutions), and including a multiplier based on the number of identified spectral lines and any multipliers from classification labels (Section 3.3.4), where they are strongly consistent or inconsistent. It is defined as follows:

$$\mathrm{rs} = \left(\sum_i^n \mathrm{ls}_i\right) \times \min\left(1, \frac{1}{2}(n^2 - n)\right) \times b, \quad (8)$$

where rs is the solution raw score, ls is a line score of an included spectral line, n is the total number of spectral lines included in the solution not counting the anchor line, and b is any multiplier from the object classification label logic (typically 0.25–2.0).

The raw score is normalized to produce the fractional score by dividing it by the sum of the raw scores of all redshift solutions.

Lastly, a scale score is produced from the weighted sum of the probability that the solution is comprised of noise, the raw score, and the fractional score as follows:

$$\mathrm{ss} = \left(1 - \prod_i^n P(\mathrm{noise})_i\right) \times w_{\mathrm{noise}}$$
$$+ \min(1.0, \mathrm{rs}/F) \times w_{\mathrm{raw}}$$
$$+ \mathrm{fs} \times w_{\mathrm{frac}}, \quad (9)$$

where ss is the scale score, $P(\mathrm{noise})_i$ is the probability that the included spectral line is noise (Section 3.1.5), w_{noise} is the weight for this first term (by default, 0.40), rs is the raw score from Equation (8), F is the configured raw score scale factor (by default, 50.0), w_{raw} is the weight for this second term (by default, 0.50), fs is the fractional score, and w_{frac} is the weight of this third term (by default, 0.10).

3.4. P(LAE)/P(O II)

P(LAE)/*P*(O II) (sometimes as PLAE/POII in other documentation) represents the ratio of the relative probability that, given a set of measured characteristics, an emission line is Lyα (representing an LAE) rather than [O II]. These probabilities are based on the number of galaxies expected at the volume sampled by the redshift slices assuming the emission line is either Lyα or [O II] given the measured line flux and equivalent width. The expected number of galaxies derives from the equivalent width distributions of Lyα and [O II] conditioned on the luminosity functions found in Gronwall et al. (2014), Ciardullo et al. (2013) respectively, interpolated or extrapolated as needed (see also Leung et al. 2017, Figure 2).

This is an improvement on the commonly used 20 Å equivalent width cut (Gronwall et al. 2007; Adams et al. 2011) and is based largely on the analysis of Leung et al. (2017), and using the specific translation and implementation described in Farrow et al. (2021, primarily in Section 2). ELiXer slightly updates Farrow et al. (2021) by (1) using multiple independent or semiindependent estimates of the continuum (Section 3.2), (2) combining those estimates into a single, best-fit continuum value, and (3) sampling over the uncertainties in the measured line flux and continuum estimates to generate a (68%) confidence interval around each *P*(LAE)/*P*(O II) measurement. Partly for convenience and partly as a representation of the practical limits of this method, the ratio is cropped to values between $0.001 \leqslant P(\text{LAE})/P(\text{O II}) \leqslant 1000$.

The interpretation of the *P*(LAE)/*P*(O II) value is not quite straightforward. While LAE evolution between $2 < z < 4$ appears somewhat muted (Blanc et al. 2011; Santos et al. 2021), there is more redshift evolution of the [O II] systems (Gallego et al. 2002; Park et al. 2015; Comparat et al. 2016; Saito et al. 2020; Gao & Jing 2021) for $z < 0.5$. This evolution may be underrepresented in the base *P*(LAE)/*P*(O II) code and lead to a deviation from the expectation that a ratio near 1 should be interpreted as the likelihood of the emission line being Lyα or [O II] is approximately equal. Building on the suggestion in Leung et al. (2017) of using different thresholds for the *P*(LAE)/*P*(O II) ratio at different observed wavelengths, ELiXer adopts an empirical threshold relation (Section 3.5.3).

The overall combined *P*(LAE)/*P*(O II) value and its confidence interval factor significantly in the final automated classification of the emission line. It can frequently be the most influential (and sometimes the only) metric that is used in that classification (Section 3.5).

3.5. P(LyA)

Using some of the features and/or measurements already described, along with a set of additional features described below, ELiXer synthesizes an aggregate confidence in the classification of the anchor emission line as Lyα or not-Lyα. For familiarity, this is couched in terms of a probability, labeled as *P*(Lyα) with values between 0 (definitely not Lyα) and 1 (definitely Lyα), but is not a true probability in the formal sense. *P*(Lyα) is the result of a weighted voting system where each of the features described in this section provides a vote (typically 0 or 1, but can be in between), and that vote is given a weight based on the robustness or confidence of the measurement. With specifically noted exceptions, the features that do not produce a clear preference are given zero or very

little weight. The final *P*(Lyα) value is then simply the sum over all votes multiplied by their respective weights:

$$P(\text{Ly}\alpha) = \frac{\sum_i (\text{vote}_i \times \text{weight}_i)}{\sum_i \text{weight}_i}. \qquad (10)$$

Note that the sum of the weights alone is not normalized and can exceed 1. In the relatively rare cases where the sum of all weights is less than 1, a special "uncertainty" vote is added with a value of 0.5 and a weight equal to $1 - \sum$ weights, so that the weights do sum to 1. This helps capture the uncertainty in the classification and prevents one or two votes with very low weights from being dominant.

The selection of voting criteria and the weights applied to the votes is the result of empirical analysis and trial-and-error testing and is discussed in Section 4. This is a little bit of the central limit theorem and the wisdom of the crowd, even though the votes are not entirely independent as several incorporate similar elements, and some are designed to handle edge cases not well covered by the others. No single vote is universally dominant; though each can be decisive under the right circumstances, such as the high weight of Section 3.5.2 when multiple emission lines are present or even a low weight vote from Section 3.5.5 for some moderate equivalent widths when the rest of the vote tally is near 0.5.

As a word on the notation in this section; often [O II] is used in place of "not-Lyα" as [O II] is the most common contaminant. The votes "for [O II]" are really votes for "not-Lyα." Further, the figures in this subsection all show only those assessment sample detection emission lines that are Lyα or [O II], so [O II] is equivalent to "not-Lyα."

3.5.1. Object Size Vote

In cases where a counterpart is identified and resolved in the *g*- or *r*-band imaging, the angular and physical extent of the counterpart contributes a vote. For this purpose, an object is considered resolved if the angular major diameter is greater than 1.1× the seeing FWHM. This includes artificially enlarged footprints in the imaging due to the *blooming* of bright sources that have saturated the detector. The proper physical diameter is computed assuming the redshift of [O II], as larger objects tend to be more evolved and at lower redshift. The emission line FWHM is used to help break the size degeneracy between larger, lower-*z* objects and saturated, higher-redshift sources, via the assumption that the latter are AGN with a large emission line FWHM.

The parameter thresholds are set from a manual partitioning of classifications in scatter plots of angular and physical diameter versus the observed wavelength of the anchor emission line, as shown in Figure 1. The conditions and their associated votes are summarized in Table 3. The specific limiting values of the FWHM help distinguish possible AGN with a broadened emission line, from lower-redshift galaxies. It is reasonable for an AGN to receive a vote for Lyα, but an angularly large object with a more narrow emission line is more likely an [O II] emitter. The gap between the conditions avoids a vote where it is unclear. The angular diameter threshold
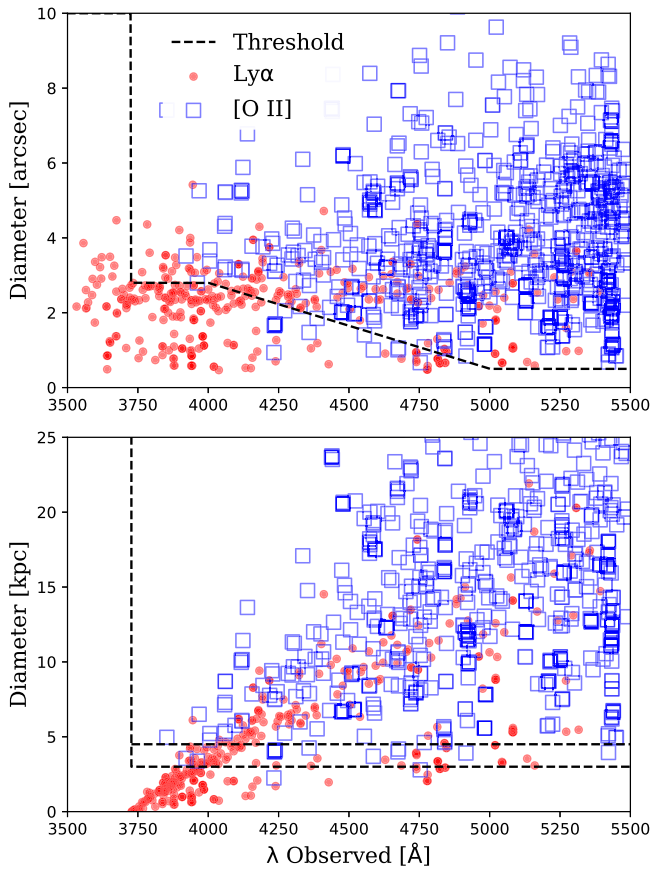
Figure 1. The separation of Lyα from [O II] in the assessment sample (SzAS, Section 4) based on the angular (upper panel) and physical (lower panel) diameters. Errors are ~0″.2. The dashed line corresponds to the thresholds defined in Table 3. There are no points blueward of 3727 Å in the lower figure because the physical diameter is computed based on the assumption that the emission line is [O II]. The lower panel is cropped to a maximum of 25 kpc for readability and shows two horizontal thresholds at 3.0 and 4.5 kpc, corresponding to the first two conditions in Table 3. This generates a vote for 70% of the SzAS with a 4% contamination of Lyα in those votes.

Table 3
Angular and Physical Diameter Votes

Condition	Vote	Weight
$d_p < 3.0$ kpc or $\theta < \theta_\lambda$	1.0	0.25
$d_p < 4.5$ kpc	1.0	0.10
$\theta < 2{.}''5$ and FWHM > 1000 km s^{-1}	1.0	0.25
$\theta > 2{.}''5$ and FWHM < 800 km s^{-1}	0.0	0.25

Note. Summary of angular and physical size votes. The conditions are ordered such that the logical evaluation results in at most one unique vote. If no conditions are met, there is no vote. d_p is the proper diameter in kiloparsecs. θ is the angular diameter in arcseconds. θ_λ is the minimum expected angular size for an [O II] galaxy for the observed anchor emission line wavelength. FWHM refers to the emission line.

(in arcseconds), θ_λ, is a piecewise linear function:

$$\theta_\lambda = \begin{cases} 2.8, & 3727\,\text{Å} < \lambda \leqslant 4000\,\text{Å} \\ -0.0018\lambda + 10.0, & 4000\,\text{Å} < \lambda \leqslant 5000\,\text{Å} \\ 1.0, & \lambda > 5000\,\text{Å}. \end{cases} \quad (11)$$

The object size criteria results in a cast vote for 70% of the SzAS (down-selected to only contain Lyα and [O II]), where
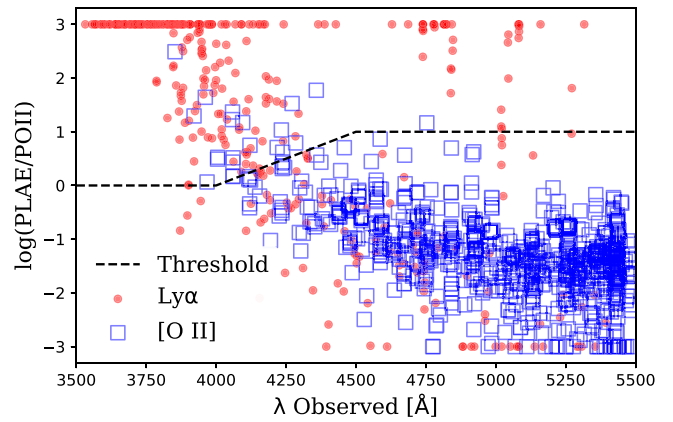
Figure 2. $P(\text{LAE})/P(\text{O II})$ distribution (clipped to $10^{\pm3}$) in the assessment sample (SzAS, Section 4) shown without the 68% confidence intervals (Section 3.5.3). The dashed line is the midpoint of the segregation threshold (Equations (15)–(18)) with points above the line receiving a vote for Lyα and those below for not-Lyα with weights based on the distance from the threshold. This vote has a 4% contamination rate of Lyα by [O II] in the SzAS.

the separation of [O II] from Lyα is effective, with a Lyα contamination rate of 4% in those votes.

3.5.2. Multiline Redshift Solutions Votes

This criterion can generate multiple votes, one for each potential redshift solution (Section 3.3) based on the positions and fluxes of the fitted spectral lines. There must be two or more found spectral lines, with the scores based largely on the number of lines and their strengths (see Section 3.3.5). However, as is shown later in this subsection, the solutions incorporating three or more lines receive an increased voting weight. At most, there will be a single Lyα (1.0) vote if there is a solution that supports the classification of the anchor line as Lyα. All other redshift solutions necessarily require the anchor line to be something other than Lyα, and therefore cast a not-Lyα (0.0) vote. The weight each vote receives depends on the scaled solution score assigned multiplied through a sigmoid:

$$w_0 = \text{ss}/(1 + \exp(0.75m - \text{rs})) \quad (12)$$

where w_0 is the initial voting weight, ss is the redshift solution scale score (Equation (9)), m is the minimum acceptable score (25, by default), and rs is the redshift solution raw score (Equation (8)).

An additional multiplier is applied for exceptionally strong redshift solutions with 3 or more contributing spectral lines:

$$w = w_0 \times \min(\text{rs}/m, 10) \quad (13)$$

where w is the modified voting weight, w_0 is the original weight (Equation (12)), rs is the raw solution score, and m is the minimum acceptable score. This multiplier is always greater than 1 since, by definition, a qualifying redshift solution must have a raw solution score greater than the minimum acceptable value. The maximum value of w is limited to $10\times$ the original number, but that allows this vote to dominate with a high confidence redshift solution comprised of multiple, strong spectral lines.

This criteria does not often trigger a vote, casting one for only 7% of the SzAS, down-selected to only contain Lyα and [O II], and 12% for the entire SzAS, but has no contamination of Lyα for those votes. Due to the bright skew in SzAS (see

Sections 4 and 5.1), this voting rate is exaggerated and is only cast for 2% of the $g > 22$ detections in HETDEX.

3.5.3. P(LAE)/P(O II) Vote

As most HETDEX detections are faint, single emission lines, the above criteria rarely produce strong redshift solutions, and the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ computation is often the most significant vote. The value (0 or 1) of the vote depends on which side of a wavelength-dependent midpoint the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ ratio falls, and the weight of the vote increases with the distance of the ratio from that midpoint. The midpoint value, which separates the [O II] (0) and Lyα (1) vote, is a modification of the binary condition suggested in Leung et al. (2017),

$$\mu = \begin{cases} 1.38, & \lambda \leqslant 4255 \text{ Å} \\ 10.3, & \lambda > 4255 \text{ Å} \end{cases} \qquad (14)$$

and is defined as follows:

$$\mu = \begin{cases} 1.0, & \lambda \leqslant 4000 \text{ Å} \\ 0.018\lambda - 71, & 4000 \text{ Å} < \lambda \leqslant 4500 \text{ Å} \\ 10.0, & \lambda > 4500 \text{ Å} \end{cases} \qquad (15)$$

where μ is the midpoint or vote threshold, and λ is the wavelength of the anchor emission line. Ratios nearer the midpoint suggest an increasingly equal likelihood that the source emission line is [O II] or Lyα and, as such, add little evidence for a classification. This is reflected in a low voting weight (w) built from a Gaussian,

$$w = 1 - \exp\left(-\left(\frac{P - \mu}{\sqrt{2}\ \sigma}\right)^2\right) \times (1 - i); \qquad (16)$$

where

$$P = \begin{cases} P(\mathrm{LAE})/P(\mathrm{O\,II}), & \text{for } P(\mathrm{LAE})/P(\mathrm{O\,II}) \geqslant 1 \\ P(\mathrm{O\,II})/P(\mathrm{LAE}), & \text{for } P(\mathrm{LAE})/P(\mathrm{O\,II}) < 1 \end{cases}, \qquad (17)$$

and μ is the midpoint, and σ is the usual Gaussian width (here set to 5.0, which is tuned by hand to give balanced voting weights). The parameter i is an ersatz standard deviation from the scaled 68% confidence interval around the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ (Section 3.4) and is defined as follows:

$$i = \frac{1}{2} \times \left(\frac{U}{U + 1} - \frac{L}{L + 1}\right) \qquad (18)$$

where U is the upper bound of the confidence interval, and L is the lower bound. As the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ ratio moves farther from the midpoint in either direction, the weight of the vote increases and rapidly asymptotes to 1.

Alone, the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ vote is effective (Figure 2), with a 4% Lyα contamination rate (by [O II]) in the SzAS (down-selected to only contain Lyα and [O II]), voting 90% of the time. As with the other equivalent-width-based votes, though, it struggles to identify Lyα emission lines when originating from non-LAE (i.e., low-EW Lyα emitting galaxies; see also Section 5.3). As the $P(\mathrm{LAE})/P(\mathrm{O\,II})$ computation includes the volumes sampled by the two assumed redshifts, it can become a less effective discriminator as the observed wavelengths approach the rest wavelength of [O II] and that volume shrinks (Section 3.4; and Leung et al. 2017; Farrow et al. 2021). The other votes, including two more based partly
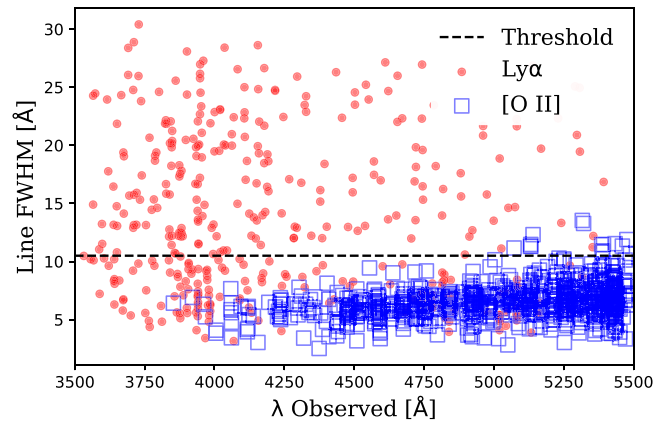

Figure 3. Lyα and [O II] separation in the assessment sample (SzAS, Section 4) based on the emission line FWHM. The data points are shown without their uncertainties ($\sim$14%). The horizontal dashed line represents the minimum threshold to receive a vote for Lyα as described by Equation (19).

on the emission line equivalent width, Section 3.5.5 in particular, help compensate.

3.5.4. Line FWHM Vote

This is logically one of the simplest votes. If the emission line FWHM is larger than 10.5 Å, as seen in Figure 3, the line receives a Lyα vote (1) with a weight as high as 1.0 using

$$w = \min(\mathrm{FWHM}/10.5 - 1.0, 1.0), \qquad (19)$$

where w is the assigned weight of the line, and FWHM is line's fitted full width at half maximum. As the contamination rate decreases with larger FWHM thresholds, the voting weight increases. If the lower uncertainty bound of the fitted FWHM, here defined as the fitted FWHM minus the uncertainty derived from standard error propagation, exceeds a configurable minimum (15.3 Å by default), the vote weight is set to the 1.0 maximum value, as [O II] emission lines are rarely that broad. Also, as a consequence of the increasing FWHM threshold, these higher weighted votes tend to favor AGN selection and thereby help reduce the confusion caused by lower AGN emission line equivalent widths. In short, it helps improve the recovery of Lyα (and decrease the misclassification as [O II]) from AGN that can fail the other voting criteria based on equivalent width (Sections 3.5.3, 3.5.5), bandpass magnitude (Section 3.5.7), and angular size (Section 3.5.1).

This criteria casts a vote for 23% of the down-sampled SzAS (containing only Lyα and [O II]) with a total Lyα contamination of 11%. This drops to 3% when considering votes with weights above 0.3 (received by 18% of the down-selected SzAS) and is contamination free for votes with weights above 0.7 (received by 12% of the down-selected SzAS).

We note that while this particular vote is a good discriminator against [O II], it can confuse Lyα with other broad AGN lines, such as C III] or C IV. We largely address this issue using multiline redshift solutions (Section 3.3) and clustering (Section 3.7).

3.5.5. Simplified Equivalent Width Vote

This vote is somewhat redundant with the full $P(\mathrm{LAE})/P$(O II) vote (Section 3.5.3), but does not consider the redshift-based population distributions or observed wavelength
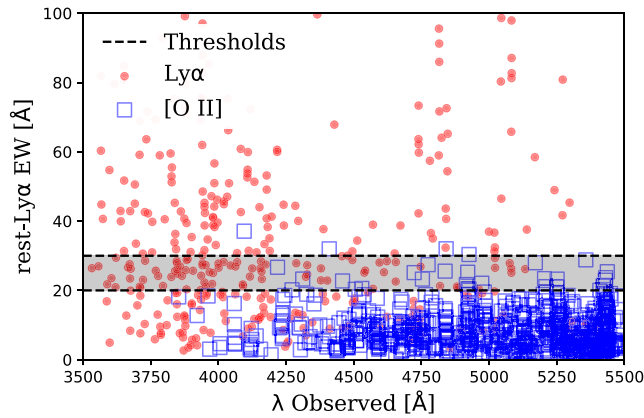
Figure 4. Simplified rest-Lyα equivalent width vote applied to the assessment sample (SzAS, Section 4). The figure is cropped to a maximum EW of 100 Å for readability and plotted without the $\sim$16% errors. The SzAS contains no spectroscopically confirmed [O II] emission lines with rest-Lyα EW > 80 Å. Detections with EWs falling in the gray shaded region between 20 and 30 Å receive no vote while those above receive a Lyα vote and those below an [O II] vote with the weight of the vote modulated by the distance to the nearest threshold (Section 3.5.5).

variations. It slightly moderates the $P(\text{LAE})/P(\text{O II})$ vote and can help push away from an uncertain classification where the $P(\text{LAE})/P(\text{O II})$ vote has a low weight. It can also push toward an uncertain classification if the $P(\text{LAE})/P(\text{O II})$ vote and this vote have similar weights, but different votes, allowing other voting criteria to have more influence. These two votes agree 95% of the time, and this simplified equivalent width vote is only important in these boundary cases.

This simplified vote uses $\text{EW}_{\text{Ly}\alpha}$, which is defined by the Gaussian fitted line flux (Section 3.1.3) and the combined continuum estimate (Section 3.2). For $\text{EW}_{\text{Ly}\alpha}$ much greater or much less than 20 Å, this reinforces the $P(\text{LAE})/P(\text{O II})$ vote and helps nudge the solution away from the $P(\text{LAE})/P(\text{O II})$ midpoint. If the $\text{EW}_{\text{Ly}\alpha}$ is greater than 30 Å, then the vote is for Lyα (1); if the $\text{EW}_{\text{Ly}\alpha}$ is less than 20 Å, the vote is for [O II] (0). All other $\text{EW}_{\text{Ly}\alpha}$ values do not generate a vote.

The assigned voting weights are based on the $\text{EW}_{\text{Ly}\alpha}$ lower ($\text{EW}_{\text{Ly}\alpha}^-$) and upper ($\text{EW}_{\text{Ly}\alpha}^+$) bounds and increase with conditions where the contamination is reduced. The maximum weight is limited to 0.5 so that the $P(\text{LAE})/P(\text{O II})$ vote is dominant when both votes approach their maximum weights.

In the pro-Lyα case, the weight is either 0 or between 0.1 and 0.5 as follows:

$$w = \begin{cases} 0.0, & r^- \leqslant 0.0 \\ 0.1, & 0.0 < r^- < 1.0 \\ \max(0.1, \min(0.5, r^- - 1.0)) & r^- \geqslant 1.0 \end{cases} \quad (20)$$

where w is the assigned weight, and $r^- = \frac{1}{25} \times \text{EW}_{\text{Ly}\alpha}^-$.

In the pro-[O II] case, the weight is between 0.1 and 0.5 as follows:

$$w = \begin{cases} 0.1, & r^+ < 1.0 \\ \min(0.5, \max(0.1, f)), & r^+ \geqslant 1.0 \end{cases}; \quad (21)$$

$$f = -0.04 \times \text{EW}_{\text{Ly}\alpha}^+ + 0.9 \quad (22)$$

where w is the assigned weight, and $r^+ = 20 / \text{EW}_{\text{Ly}\alpha}^+$.

Figure 4 shows the Lyα and [O II] SzAS detections with rest-Lyα EW less than 100 Å (this includes all SzAS [O II] emission lines) with the voting thresholds marked.

This criteria votes in 80% of the down-selected SzAS (containing only Lyα and [O II]) with a Lyα contamination rate of 2%. Superficially, this is superior Lyα/[O II] segregation compared to the $P(\text{LAE})/P(\text{O II})$ vote (Section 3.5.3) but, by design, avoids voting in the difficult EW transition region (shaded region in Figure 4).

3.5.6. Catalog Photometric Redshift Vote

The photometric redshift fits from the various included catalogs (Section 2) are often too broad to confidently pin down a tight redshift constraint. However, they can be sufficient to distinguish between low-z ($z \lesssim 0.7$) and high-z ($1.7 \lesssim z \lesssim 3.7$) objects and thus help separate [O II] from Lyα. If there are any photometric redshifts for an HETDEX detection, this vote simply takes the arithmetic mean of all phot-z measurements of the matched catalog counterpart from all contributing catalogs and compares it to the low-z and high-z ranges quoted above. If the mean falls within either range and is within a redshift distance of 0.5 of [O II] or Lyα respectively, then the corresponding vote is cast with a weight of 0.5. If the mean falls outside of both ranges or if the redshift separation between the mean and an assumption of [O II] or Lyα is greater than 0.5, then no vote is cast.

Only $\sim$1.5% of HDR3 sources have at least one phot-z catalog counterpart match, so this vote rarely contributes to the $P(\text{Ly}\alpha)$ logic. For the SzAS testing, because the contributions from catalog phot-z and spec-z are necessarily turned off, this vote is never cast.

3.5.7. Apparent Magnitude and Equivalent Width Vote

This vote is largely predicated on the observation that the HETDEX LAEs tend to be fainter than [O II] galaxies. However, there certainly exist bright LAEs (including AGN) and faint [O II] galaxies, so the $\text{EW}_{\text{Ly}\alpha}$ is also incorporated into the decision to moderate it.

The apparent magnitude used in this vote is the g-band magnitude derived from the HETDEX spectrum (Section 3.2), which has a limiting magnitude of $\sim$25$_{AB}$. The magnitude threshold between votes for [O II] and for Lyα is defined by a pair of lines whose parameters are set to optimize the segregation of those two samples. The objects with g magnitudes fainter than the upper line of Figure 5 are more likely to be Lyα, while those brighter than the lower line of Figure 5 are more likely to be [O II]. The classification of objects, defined by their SzAS spectroscopic redshifts, lying between these two regimes is uncertain. The optimization over the slope and intercept parameters of these lines was performed using a simple grid search that maximizes the Lyα accuracy in one case and the [O II] accuracy in the other. While an MCMC fit could be more precise, given the uncertainties in the data features and the desire to avoid overfitting to the specific test set, the grid search is preferred. The accuracy is defined as follows:

$$\text{accuracy} = 1 - \frac{\xi + \epsilon}{\Omega} \quad (23)$$

where ξ is the number of *true* Lyα ([O II]) detections (here as the spec-z counterparts in the SzAS test sample), which are not identified by the selection; ϵ is the number of incorrectly classified Lyα ([O II]) detections; and Ω is the total number of
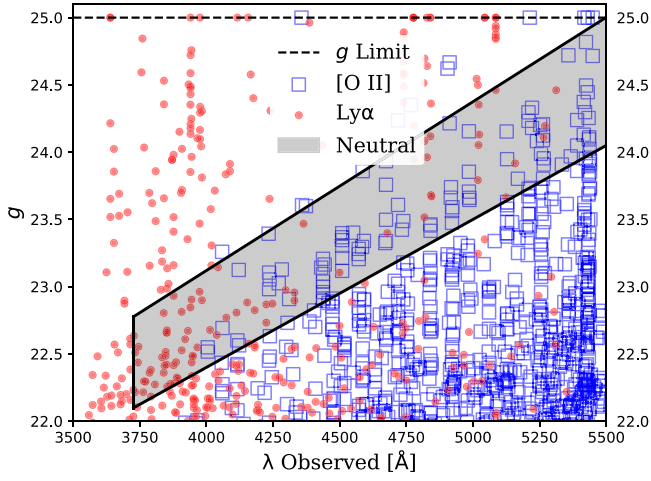
Figure 5. The apparent magnitude (error ∼0.1) and equivalent width vote, by itself, is highly effective at segregating Lyα from [O II] against the assessment sample here (SzAS, Section 4). The *neutral* region is defined by the lines of Equations (24) and (25) as the lower and upper bounds respectively, and extends from 3727 Å to the red edge of the HETDEX spectral window. Lyα emitters represent 97% of the down-selected SzAS above the *neutral* region, 44% inside the *neutral* region, and 14% below the *neutral* region.

Lyα ([O II]) classified detections. Here, *true* is assumed as the catalog based spectroscopic redshifts (Section 4).

The two lines are defined as follows:

$$g_T^+ = 1.10 \times 10^{-3}\lambda + 18.0, \text{ (lower line, Fig. 5);} \quad (24)$$

$$g_T^- = 1.26 \times 10^{-3}\lambda + 18.1, \text{ (upper line, Fig. 5)} \quad (25)$$

where g_T^+ is the faint magnitude threshold, g_T^- is the bright magnitude threshold, and λ is the wavelength (Å) of the anchor emission line.

We also define upper (faint) and lower (bright) bounds for the measured g magnitude of each detection (g^+ and g^-, respectively) based on the propagated errors of the HETDEX spectroscopically measured g-band magnitude.

The votes and their weights for this criterion are summarized in Table 4 with Figure 5 showing the segregation of Lyα and [O II] with Equations (24) and (25). As the g magnitude becomes brighter, the voting weights for Lyα decrease, and those for [O II] increase. The large anchor line $EW_{Ly\alpha}$ favors Lyα, and the small $EW_{Ly\alpha}$ favors [O II]. With the exception of spectra associated with objects having faint g magnitudes, those spectra with anchor line $EW_{Ly\alpha}$ (with error) between 15 Å and 30 Å receive no vote either way. Contamination of Lyα by [O II] for the down-selected SzAS is low with Lyα comprising 97% of the detections above the *neutral* region in Figure 5. Conversely, Lyα represents only 44% within the *neutral* region, where no vote is cast, and 14% below it, where the vote is cast for [O II].

3.5.8. Disqualifications

Disqualification conditions are a set of special classifications and data integrity issues that can either contribute additional weighted votes against a Lyα classification or, in extreme cases, completely override the $P(Ly\alpha)$ results.

1. *Meteor*. If the detections have a possible classification as a meteor (Section 3.3.4), a vote against Lyα is added with a weight equal to the strength of the meteor classification

Table 4
Apparent Magnitude and EW Votes

Condition	Vote	Weight
$g^- > g_T^+$	1.0	0.50
$g_T^- < g^- < g_T^+ < g^+$ and $EW^- > 80$	1.0	0.50
$g_T^- < g^- < g_T^+ < g^+$ and $EW^- > 30$	1.0	0.30
$g_T^- < g^- < g_T^+ < g^+$ and $EW^+ \leqslant 15$	0.0	0.25
$g^- < g_T^- < g^+ < g_T^+$ and $EW^- > 80$	1.0	0.30
$g^- < g_T^- < g^+ < g_T^+$ and $EW^- > 30$	1.0	0.15
$g^- < g_T^- < g^+ < g_T^+$ and $EW^+ \leqslant 15$	0.0	0.40
$g^+ < g_T^-$ and $EW^- > 80$	1.0	0.25
$g^+ < g_T^-$ and $EW^- > 30$	1.0	0.10
$g^+ < g_T^-$ and $EW^+ \leqslant 15$	0.0	0.50

Note. Summary of apparent magnitude and equivalent width votes. The conditions are ordered such that the logical evaluation results in at most one unique vote. If no conditions are met, there is no vote. The apparent g magnitude becomes brighter moving down the table. g_T^+ is the upper (faint) g threshold as a function of λ. g_T^- is the lower (bright) g threshold as a function of λ. g^+ is the upper bound (faint) g for the detection. g^- is the lower bound (bright) g for the detection. EW^+ is the upper bound rest-frame EW in Å, assuming Lyα. EW^- is the lower bound rest-frame EW in Å, assuming Lyα.

(0.0–5.0). Given its potentially large weight, this vote can be dominant. Regardless of the final result of the vote, the "meteor" label is attached to the detection.

2. *Bad pixel flat*. If a bad pixel flat is indicated by pixel-to-pixel variations or pixel flux values outside the acceptable range for an emission line on that part of the CCD, then the emission line may be entirely due to, or at least enhanced by, this artifact. The detection will thus receive a vote against Lyα with a weight equal to 1.0 plus the sum of the relative weights of those fibers contributing to the spectrum that have a bad pixel flat. The total weight for this vote is between 1.0 and 2.0. However, if the sum of the fiber weights exceeds a threshold, 0.50 by default, the entire $P(Ly\alpha)$ vote is disqualified. Independent of the vote, the bad pixel flat flag is associated with the detection and shown on the ELiXer report.

3. *Duplicate fibers*. If duplicated fibers (identified by repeated fiber identifiers or identical flux and error data vectors) appear in the detection spectra, the $P(Ly\alpha)$ vote is disqualified. This is an indication of a data reduction problem.

4. *Grossly negative spectrum*. If less than 10% of the wavelength bins contain nonnegative integrated flux values, the spectrum is considered *grossly negative* and suggests some issue in the reduction. In this case, the detection and the $P(Ly\alpha)$ vote is disqualified.

5. *Poor observation*. If the seeing FWHM is worse than a threshold (3″ by default) or the throughput response, as defined by Gebhardt et al. (2021), is less than a threshold (0.08, by default), the input observation is considered too poor to make a meaningful classification attempt, and the vote is disqualified.

6. *Bad dither norm*. If the dither-to-dither normalization (Gebhardt et al. 2021) for the detection is above a threshold (3.0× by default), a potentially severe

observation or reduction issue is indicated, and the vote is disqualified.

3.6. Best-z and Q(z)

Unless there is a serious error or a disqualification (Section 3.5.8), ELiXer assigns a single, best guess redshift, *best-z*, along with a quality score, $Q(z)$, as an indication of the confidence in that redshift. The assignment of the best-z incorporates all prior information and analysis including the $P(Ly\alpha)$, catalog spec-z and phot-z, and any multiline redshift solutions (Section 3.3). The $Q(z)$ value takes on a continuous value between 0 and 1, with 1 meaning *full confidence* and 0 meaning *no confidence* (i.e., the redshift is effectively a guess). Where the $P(Ly\alpha)$ analysis is limited only to a determination as to whether the emission line is Lyα the best-z logic attempts to fully specify the redshift. In the ideal scenario, there are multiple high-S/N emission lines within the HETDEX spectrum, each corresponding to a known line at a consistent redshift. In such a case, the best-z is clear and the corresponding $Q(z)$ is 1.0. Such objects are rather rare, but they do define the starting benchmark.

The best-z is set as (1) the redshift from a qualified multiline spec-z solution, (2) the Lyα redshift when there is no spec-z solution but $P(Ly\alpha)$ favors Lyα, or typically, (3) the [O II] redshift. In the last case, the redshift can be set to C III] or Mg II when the line is broad and occurs at a wavelength within the HETDEX spectral window where no other strong feature is expected to be found.

The $Q(z)$ confidence value is set based on the best-z selection condition. It is primarily a function of the $P(Ly\alpha)$ value and the multiline solution score. $Q(z)$ is maximized by $P(Ly\alpha)$ when $P(Ly\alpha)$ is near 0 or 1 and minimized when $P(Ly\alpha)$ is near 0.5. The effect of the multiline solution score, on the other hand, is a monotonic increase with the multiline solution score. $Q(z)$ may also have penalties and caps imposed on it based on specific circumstances and flags, such as the detection being near a spatially extended, bright object or if the various continuum estimates (Section 3.2) disagree. If the multiline solution and $P(Ly\alpha)$ agree, the $Q(z)$ score increases; if the two measures disagree, the $Q(z)$ score is decreased based on the relative difference between the multiline solution and $P(Ly\alpha)$ strengths. The selection logic and $Q(z)$ assignment is summarized in Table 5.

Since the majority of HETDEX objects are faint, with a single detected emission line, most (~80%) receive a $Q(z)$ score less than 0.5 with ~35% in the lowest $Q(z)$ bin (0–0.1). These are still usually correctly classified as is shown in Sections 4.4 and 5, but rely on less evidence and thus have a low $Q(z)$ value.

3.7. Clustering–Neighbor Redshift Matching

In the low-surface brightness outer regions of spatially resolved galaxies, HETDEX detections with low, PSF-weighted line fluxes (commonly arising from faint H II regions and PNe) may be incorrectly classified by ELiXer as Lyα. To address this issue, ELiXer can optionally compare a detection against other nearby HETDEX detections and look for consistencies. When invoked, ELiXer examines all HETDEX emission line detections within $15''$ (by default) of the current detection under consideration, and tests for g-band magnitudes brighter than 23_{AB} with matching observed emission line(s) of

Table 5
Best-z, $Q(z)$ Summary

Condition	Best-z	$Q(z)$
Strong, multiline spec-z solution consistent with $P(Ly\alpha)$	multiline spec-z	4–5★
Strong, multiline spec-z solution not consistent with $P(Ly\alpha)$	multiline spec-z	0–3★
Weak, multiline spec-z solution consistent with $P(Ly\alpha)$	multiline spec-z	2–4★
Weak, multiline spec-z solution not consistent with $P(Ly\alpha)$	multiline spec-z	1–3★
$P(Ly\alpha)$ only, $\gtrsim 0.7$	Lyα	3–4★
$P(Ly\alpha)$ only, $\gtrsim 0.5$	Lyα	0–2★
$P(Ly\alpha) \lesssim 0.5$ with single, broad emission line	[O II] Mg II, C III]	0–1★
$P(Ly\alpha)$ only, $\lesssim 0.5$	[O II]	0–1★
$P(Ly\alpha)$ only, $\lesssim 0.3$	[O II]	0–2★

Note. Summary of the best-z and $Q(z)$ logic. Specific values (0.0–1.0) of the $Q(z)$ are not shown as they depend on details omitted, but are expressed as these qualitative descriptors: 5★ (~1.0), 4★ (~0.80), 3★ (~0.50), 2★ (~0.35), 1★ (~0.25), 0★ (~0).

higher line score (Section 3.1.4). The presumption, which is borne out in testing, is that the brighter, higher-scoring detections are (1) better centered on the object and (2) more likely to receive the correct classification. The requirement to match the observed emission line wavelength(s) in addition to the on-sky proximity helps preserve the classification of background objects with lines of sight passing near the brighter foreground source. When more than one match is found, the highest-scoring redshift solution is selected, and if the selected object is brighter and higher scoring than the current detection's solution, that neighbor's classification is used as a replacement. In other words, the faint, low-scoring detections can be assigned the more secure redshift of an immediately adjacent, brighter, higher-scoring *neighbor* detection when they share matching observed-frame emission lines and are assumed to represent different detections of the same object. When this update is made, the altered detection is marked with a flag and the detection ID number of the matching neighbor detection.

This clustering has a relatively small effect, modifying less than 0.5% of all HETDEX emission line detections. The algorithm does not link nor otherwise combine the individual detections; all detections remain uniquely reported.

4. Testing and Results

All the effort made toward classification is effectively meaningless without appropriate testing and a selection of a reasonable spec-z assessment sample (SzAS) against which to test. As HETDEX is a large and unique survey with no preselection of targets, it is impossible to collect an overlapping observational data set of known redshifts of even remotely similar size (in terms of numbers of unique astrophysical objects) and continuum depth. Beyond polling experts for classifications based on visual inspection, and comparing ELiXer results against those of simulated objects, the best we

can do is match HETDEX sources against spectroscopic redshift catalogs produced by other surveys.

The assessment sample for this work is a composite of matched HETDEX detections from the public, archival catalogs described in Section 2 and in Mentuch Cooper (2023). In all cases, these are spectroscopic redshifts only; no photo$z$ estimates are used in this assessment sample. For the catalog-provided redshifts, the source matching to HETDEX is based on sky position and apparent magnitude. The catalog source position must be inside or within $0\rlap{.}''5$ of the edge of the SEP aperture associated with the HETDEX detection if an aperture match is made (Section 2.2), or within $0\rlap{.}''75$ of the HETDEX position if the object is fainter than $g = 24.5$, and no SEP aperture is matched. The catalog-matched spectroscopic redshifts are accepted as true.

The assessment sample is down-selected to only those detections fainter than $g = 22$ with redshifts that match any of the emission lines in Table 2 to within $\pm 4\,\text{Å}$. Although the magnitude distribution still significantly skews to brighter objects, this filtering helps refine the selection to better align with the more common, fainter HETDEX detections. The result is a data set consisting of 834 [O II] emission lines, 384 Lyα lines, and 402 *other* lines, including C IV, C III], Mg II, and Hβ as reported in the SzAS. Each redshift corresponds to a unique HETDEX detection; however, these are not necessarily unique galaxies. For brighter, extended galaxies there can be more than one overlapping HETDEX emission line detection, and where there are multiple observations covering the same position, the same galaxy may be detected more than once. Because ELiXer operates on each HETDEX detection individually, this is as intended.

4.1. Definitions

For the remainder of this work, we make the following definitions:

1. *Accuracy*. This is the number of agreements between the ELiXer assigned classification and the SzAS classification divided by the number of ELiXer detections of that classification. A match is counted if the rest-frame wavelengths from the HETDEX observed wavelength, and the SzAS and ELiXer assigned redshifts agree within $\pm 4\,\text{Å}$.
2. *Recovery*. This is a fraction roughly equivalent to completeness, but with no correction made for survey biases. Here we refer to the number of detections of a particular emission line, identified by the ELiXer software, that are matched 1:1 to that of the SzAS divided by the number of those emission lines in the SzAS.
3. *Contamination*. The fraction of detections within some defined range that are incorrectly classified. This may be further refined to the fraction of misclassifications by a particular emission line. For example, we will discuss the contamination in the Lyα sample by [O II] as a function of g magnitude.

Accuracy can be slightly underreported for broad, noisy lines where the fitted line center can be offset from the true center and where winds and radiative transfer effects can create a significant velocity offset from the systemic redshift. The $\pm 4\,\text{Å}$ allowance covers all but the most extreme cases, so the impact is minimal. Accuracy and contamination are direct inverses, and for any given emission line, they necessarily sum to unity.

Accuracy and recovery are similar, but differ by the base divisor. For the recovery of detections, any contamination of one emission line comes at the direct cost to the recovery of another emission line. Conversely, the recovery counts of an emission line are also 1 minus the sum of the contaminations of all other emission line types. Notice that the relationship does not directly hold for recovery and contamination *rates*, as each of those rates have different divisors.

4.2. Calibration

Testing and calibration are combined in a highly iterative process. ELiXer is executed on the detections of the test data set, but with catalog matching spec-z and phot-z turned off. That is, for the test runs, ELiXer does not include or consider the catalog-reported spectroscopic redshifts that would, in a standard run, factor into the classification. The ELiXer output, specifically the $P(\text{Ly}\alpha)$ values and the redshifts, is then compared to the test sample and checked for contamination, recovery, and accuracy. Disagreements between the ELiXer results and the assessment sample are examined, and manual adjustments to the individual votes and voting weights (Section 3, and Section 3.5 in particular) are made as warranted. Considerations against overturning and potentially incorrect test sample redshifts are addressed with deliberately loose fitting, low-order segmentation thresholds and by varying the composition of the test sample by creating random and targeted (in apparent magnitude, line FWHM, observation field, etc.) subsets. The process is repeated until there is good agreement (generally, matching 90%–95% or better) between the ELiXer assigned redshifts and the test sample redshifts. With the focus on $P(\text{Ly}\alpha)$ as the primary classification metric and with its flexible threshold selection, what constitutes *good* agreement is somewhat subjective but is also highly adaptable to the specific scientific needs. For example, the stacking of spectra to measure Lyman Continuum in Davis et al. (2021) is very sensitive to contamination but does not specifically require a highly complete sample, and so it utilizes a $P(\text{Ly}\alpha)$ selection of 0.8 and greater. On the other hand, the $H(z)$ and $D_A(z)$ precision goals for the primary HETDEX science are less sensitive to contamination but need to be largely complete (Farrow et al. 2021; Gebhardt et al. 2021), and a $P(\text{Ly}\alpha)$ threshold of 0.5, or even lower, is more appropriate.

4.3. Additional Testing

To supplement the catalog spec-z testing, several other testing and feedback efforts are actively used. Although the mechanics vary, all provide checks on the ELiXer classifications with targeted detection subsets. As with the SzAS, the detections where these alternate methods and ELiXer disagree are manually inspected, and adjustments to the ELiXer classification algorithm(s) are made as warranted.

These supplementary efforts fall into two categories. The first are automated machine-learning classifiers, both supervised and (sometimes) unsupervised. These are all in early development and explore various classification frameworks, with both T-distributed stochastic neighborhood embedding (van der Maaten & Hinton 2008) and autoencoder neural network (Wang et al. 2014) techniques showing good promise.

The second category relies on manual, visual vetting. The first efforts focused on HETDEX collaboration experts and university students (after receiving training). A more recent
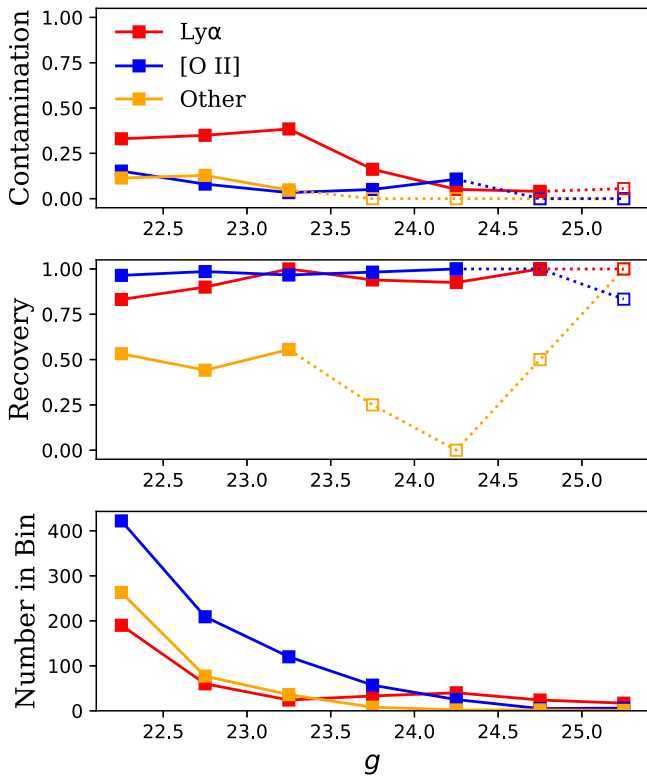
Figure 6. Performance summary of ELiXer classification and redshift assignment vs. the SzAS in *g*-magnitude bins. ELiXer does very well with Lyα and [O II], as intended, but struggles with the *other* lines, such as C IV λ1550, C III] λ1909, and Mg II λ2800. Note that the results for the faintest bin for Lyα, the faintest 2 bins for [O II], and the faintest 4 bins for other lines, denoted with open markers and dotted lines, have too few SzAS counts to be meaningful. The high contamination rate in Lyα at brighter magnitudes is a result of the biases in the SzAS and is discussed in Section 5.

Table 6
Lyα vs. [O II] Segregation on Assessment Sample

Method	Lyα Contamination	Lyα Recovery
Lyα rest EW > 20 Å	0.084	0.708
P(LAE)/P(O II) default[a]	0.090	0.763
P(LAE)/P(O II) optimized[a]	0.056	0.724
P(LAE)/P(O II) ELiXer[b]	0.042	0.705
P(Lyα) > 0.7	0.005	0.752
P(Lyα) > 0.6	0.007	0.797
P(Lyα) > 0.5[c]	0.010	0.903
P(Lyα) > 0.4	0.027	0.926
P(Lyα) > 0.3	0.056	0.940

Notes. The cumulative performance of various methods against the SzAS down-selected to only include [O II] (834 detections) and Lyα (384 detections). This allows a fairer comparison of P(Lyα) (Section 3.5) to the first three methods, which do not consider other lines. The SzAS is biased to bright objects, with an overrepresentation of AGN, so these results do not directly translate to the larger population of HETDEX detections. An adjustment for these biases are made and discussed later in Section 5.1. Additionally, although efforts are made to avoid overfitting to the SzAS, its detections significantly contribute to the determination of the votes and weights of the P(Lyα) metric, so these results may not be as representative when considering all HETDEX detections.
[a] Leung et al. (2017).
[b] Modified P(LAE)/P(O II) optimized used in ELiXer (Section 3.5.3).
[c] Default input to best-*z* logic (Section 3.6).

science outreach effort has opened classification and general exploration to the public in a citizen science project on *Zooniverse*[25] One workflow of the *Dark Energy Explorers*[26] project (L. House et al. 2023, in preparation) tasks its citizen scientists to classify HETDEX detections as being either at low-*z* ("nearby galaxy or star") or possibly high-*z* ("distant galaxy or nothing") using a reduced ELiXer report that contains only sections of 2D fiber cutouts, single band (*g* or *r*) photometric imaging, and a Gaussian fit to the emission line. Each detection receives 15 responses with the aggregate classification reported as the mean of those responses. Even with this reduced information, these broad categories match with the ELiXer classification more than 92% of the time with L. House et al. 2023, in preparation, estimating 7.7% contamination and 90.7% recovery of high-*z* galaxies. As with the other methods, select disagreements between ELiXer and *Zooniverse* are reviewed for potential classification failures by ELiXer.

4.4. Results Summary

A comparison of the ELiXer classification and redshift assignments with those of the SzAS is summarized in Figure 6 and in Table 6. Figure 6 breaks out the contamination and recovery rates by *g* magnitude, with the counts of each type

shown as a reference in the bottom panel. When there are very few classifications of a given type, such as faint [O II] and *other* lines, the accuracy and recovery rates are not meaningful. Against the SzAS, ELiXer performs very well on Lyα and [O II] classifications, but is challenged by the *other* emission lines. As will be discussed later, the elevated contamination in the Lyα detections at bright magnitudes is a function of the biases in the SzAS as compared to the HETDEX survey.

Table 6 summarizes the cumulative performance of several different Lyα/[O II] segregation methods against the SzAS identifications of the Lyα or [O II] line. This down-selection is made so that the comparisons of the ELiXer P(Lyα) method (Section 3.5) at several selection thresholds are equitable, as 20 Å equivalent width cut and the P(LAE)/P(O II) method do not classify lines other than Lyα and [O II]. It is clear that each method is an effective classifier. Except at the extreme thresholds, the P(Lyα) methods produce the lowest contamination and highest recovery rates, with P(Lyα) > 0.5 yielding a good balance of contamination and recovery fraction. This is the default input for the ELiXer best-*z* assignment (Section 3.6). Given the biases in the SzAS for bright objects and AGN, though, these results cannot be directly applied to the whole of HETDEX. However, a correction for these biases is made and discussed later in Section 5.1. We also caution that the detections in the SzAS factor significantly in the calibration of the votes and weights of the P(Lyα) metric. Although efforts are made to avoid overfitting, these results could still be less reflective of HDR3 in general.

The contamination rate of Lyα by [O II] is effectively flat as a function of the observed wavelength of the emission line. However, the recovery rate of Lyα sources trends lower as the observed wavelength moves redward. At the blue end of the HETDEX spectral range, $\lambda_{obs} \lesssim 4200$ Å, the recovery rate is ~97%; in the middle range, $4200 \lesssim \lambda_{obs} \lesssim 4800$, the rate is ~91%; and at the red end, $4800 \lesssim \lambda_{obs}$, the rate is ~81%. This

25 https://www.zooniverse.org/
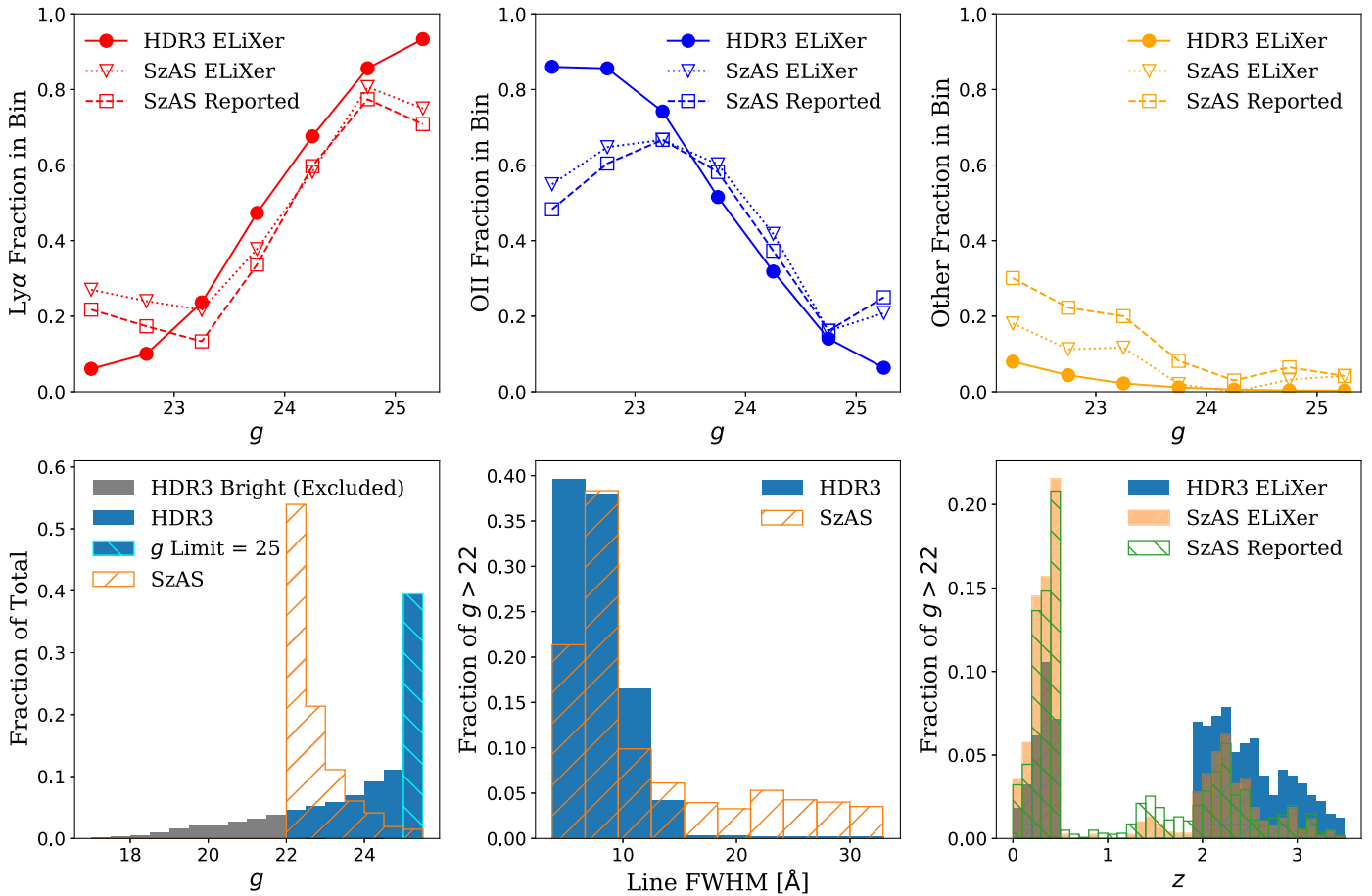26 https://www.zooniverse.org/projects/erinmc/dark-energy-explorers

Figure 7. Summary of the ~1600 emission line detections in the Spec-z Assessment Sample (SzAS) compared to the ~1.5×10^6 detections in the HETDEX Data Release 3 (HDR3). The top panels show the relative fraction of Lyα, [O II], and the other emission line detections as a function of g magnitude, as classified by ELiXer and as reported by archival spec-z measurements in the SzAS. The ELiXer reported classifications represent more of an *apples to apples* comparison, as it is clear that the SzAS is skewed toward brighter magnitudes and is significantly overabundant in the other emission line detections. The Lyα and [O II] distributions are very similar fainter than about 23.5_{AB}, but diverge at the brighter end. The lower left panel illustrates the bright bias. The lower center panel shows an excess in the SzAS for broad emission lines; although not explicitly shown here, these broad lines are predominantly Lyα, C IV 1549 Å, and C III] 1909 Å and originate from brighter, probably AGN, objects. The lower right panel echoes the overabundance of the other emission lines, showing an increase in the fraction of $1.0 \lesssim z \lesssim 2.0$ detections, likely AGN, compared to HDR3.

is an effect of larger numbers of faint [O II] emitting galaxies and fewer numbers of LAEs in their respective higher-redshift regions. These [O II] galaxies are more similar in appearance to LAEs based on several of the metrics used in ELiXer, g and r magnitudes, angular size, and even EW and line width to a lesser extent (see Sections 3.5.1, 3.5.4, 3.5.5, and 3.5.7 and their figures). The observed emission line wavelength factors in the related votes help keep the Lyα contamination rate flat and low, but at the cost of the loss of some LAEs to [O II] classifications. As shown in Table 6, this can be tuned to improve the Lyα recovery rate at the expense of a higher contamination rate as dictated by particular science needs.

5. Discussion

As can be seen from Figure 7, the sample we use for spectroscopic assessment, SzAS, is highly biased to brighter detections, somewhat biased to broader lines, and contains an overrepresentation of emission lines other than Lyα and [O II], as compared to HETDEX as a whole. At its bright end, the sample is underabundant in [O II] and overabundant in Lyα with the reverse at the faint end. Since these spectroscopic redshifts come from existing archival surveys (Section 4) and spectroscopy is historically expensive, it stands to reason that

the available spectra would favor brighter, rarer objects. An expansion of the SzAS is underway in collaboration with DESI (Martini et al. 2018; Levi et al. 2019), which will provide higher spectral resolving power ($R \sim 2000$–5000) and a redder wavelength coverage (3600–9800 Å) to selected HETDEX detections. This will increase the number of faint ($g > 25$) spectra in future assessment samples and bring their distributions more in line with HETDEX.

While not completely devoid of faint objects, the SzAS contains a smaller fraction of its detections in the faintest bins compared to the full HETDEX sample. This is not unexpected and is not a significant issue. Given the methodology of the classification, ELiXer is likely to classify anything fainter than $g \sim 25$ as an LAE in the $1.9 < z < 3.5$ redshift range. While there are certainly [O II] emission-line galaxies with $z < 0.5$ and $g > 25$, if we assume that this emission has a rest-frame equivalent width of less than 20 Å, then [O II] can be expected to be, at most, $\sim 3 \times 10^{-17}$ erg s^{-1} cm^{-2}. This maximum value is $\sim 2 \times$ fainter than the 50% flux limits for HETDEX (Gebhardt et al. 2021), making it unlikely that HETDEX would even detect an [O II] emission line from such a galaxy. Thus the reduced fraction of $g \gtrsim 25$ objects in the SzAS, compared to HDR3, is largely moot.

Nevertheless, the other biases cannot be ignored. While an uncorrected assessment sample can serve as a development test set and provide reasonable limits on the expected contamination, recovery, and accuracy rates for ELiXer classifications, a correction is needed to extrapolate to the entire HETDEX emission line sample.

5.1. Bias Correction to the Full HETDEX Catalog

Given the clearly biased distribution of the assessment sample as compared to the full HETDEX catalog, it is prudent to apply some measure of correction before extending the results from the SzAS to the full catalog. The correction chosen is relatively simplistic and, as will be shown a little later, has effectively no impact on the overall sample results.

As seen earlier, the SzAS data set is subdivided into Lyα, [O II], and other emission line detections, and each subset is binned by g magnitude from 22_{AB} to 25_{AB} in steps of 0.5, with the last bin containing all detections fainter than the 25_{AB} flux limit. The contamination (by type) for each of the three classifications is computed against the SzAS in each g bin as defined in Section 4.1.

To correct for the population biases in the SzAS compared to the full HDR3 sample, we consider the contamination rates in the SzAS to be functions of the per bin fractions of the contaminant, and the target type as classified by ELiXer. This allows us to use the same ELiXer classification rates in the full HDR3 sample as a correction to the SzAS rates. The applied correction to the SzAS values then is

$$C'_{i,j} = \frac{\left(\sum_k C_{i,j,k} \times \frac{E_{H,j,k}}{E_{S,j,k}} \times N_{H,i,k}\right)}{\sum_k N_{H,i,k}} \qquad (26)$$

where the following statements are true:

1. $C'_{i,j}$ is the corrected contamination rate of the target type i (Lyα, [O II], or other) by contaminant j, such that $i \neq j$.
2. $C_{i,j,k}$ is the directly computed contamination rate in the SzAS per g-magnitude bin, k (matching the bins in Figure 6).
3. $E_{H,i,k}$ is the ELiXer classification fraction of the target type in HDR3 per g-magnitude bin.
4. $E_{S,j,k}$ is the ELiXer classification fraction of the contamination type per g-magnitude bin in the SzAS.
5. $N_{H,i,k}$ is the number of target ELiXer classifications in HDR3 per g-magnitude bin.

An additional simple correction is also applied to help account for FPN detections caused by noise interpreted as an emission line by the HETDEX line-finding algorithm (Gebhardt et al. 2021). These are random fluctuations in the PSF-weighted spectrum from thermal electrons in the CCDs, stray photons, read noise, etc., that happen to scatter up and pass the various filtering thresholds in the line-finding code and masquerade as low S/N emission lines. They do not represent real astrophysical sources, but when interpreted as such, they map to random locations in (R.A., decl., z)-space. As the candidate emission line S/N increases toward 5.5, the incidence of these FPN rapidly approaches 0. As discussed later in Section 5.4, this has only a minimal impact on the HETDEX cosmological measurements. As an approximate correction, the ELiXer classification ratios in Equation (26) for HDR3 are modified by assuming 30% of all detections with S/N < 5.0 and 15% of all detections with $5.0 \leqslant$ S/N < 5.5 are

FPNs and simply removing those from all summed counts. Early indications are that the true FPN rates may be significantly less than this, Mentuch Cooper (2023), so we believe the assumed FPN rates are overestimates.

5.2. Performance

Figures 8 and 9 show the cumulative (bright to faint) contamination fraction of Lyα by [O II] and all *other* lines respectively, both for the SzAS and for the $g > 22$ HDR3 data set. Table 7 reports the cumulative contamination rates from those two figures (highlighted by bold typeface), provides summary information on the contamination in [O II] and the *other* lines, and gives the accuracy and recovery rates for all discussed line types. Note that the values for the SzAS corresponding to Table 6 are slightly different, because the detections for that table are down-selected to only include Lyα and [O II]. Overall, ELiXer performs extremely well in mitigating the contamination in the Lyα classification, and excels at the faint end against the primary contaminant, [O II]. This is what ELiXer is tuned to do. At brighter magnitudes, non-[O II] contaminants are more problematic; though they represent only a small fraction of the total HETDEX data set (Figure 7). For the HETDEX data releases, the final classification of these objects is assisted by the supplemental program, *Diagnose* (G. Zeimann et al. 2023, in preparation; see also Section 5.3).

The cumulative fractional contamination from [O II] has a peak between $g \sim 23.0$, and $g \sim 24.0$, where the numbers of [O II] and LAEs are most similar. The total contamination rate sits at only 1.3% for the SzAS even with the [O II] emitters outnumbering LAEs in that sample by more than 2:1. For HDR3, when corrected for the SzAS distribution bias and predicted FPNs from noise, the predicted contamination rate is 1.2%. While this already meets the HETDEX requirements, the planned ELiXer enhancements, including updated Lyα and [O II] luminosity functions for the $P(\text{LAE})/P(\text{O II})$ analysis (Section 3.4) and run-time phot-z fitting, should further decrease the contamination rate and improve overall accuracy.

The cumulative fractional contamination of Lyα from all other lines in the SzAS is substantial at 26.4%. This, however, is significantly inflated due to the overrepresentation of AGN and C III] and C IV emission lines in the SzAS (Figure 6, upper right panel). When projected onto the HDR3 distribution and corrected for the SzAS distribution bias and noise-driven FPNs, this cumulative contamination fraction falls to a predicted 0.8% for the full HDR3 data set. This is even better than the [O II] contamination. However, given the large correction from the SzAS results (Figure 9), it is prudent to estimate a worst-case contamination by these other lines by alternate means. These misclassifications in the SzAS are dominated by C III] and C IV and are characterized by bright magnitudes and large line widths—the median $g = 22.5 \pm 0.5$, and the median emission line FWHM = 22 ± 8 Å. Using these properties as a guide, we select the fraction of HDR3 detections with emission line FWHM > 14 Å and $g < 23$, yielding 5.8% of HDR3 detections, of which we assume $1/3$ are misclassified as Lyα. With 47% of detections classified as Lyα, we then estimate the worst-case contamination rate by the other lines at 4% (e.g., $\frac{1}{3} \cdot 0.058 \big/ 0.47 = 0.04$).

While this is $5 \times$ the *bias + FPN corrected* contamination rate of 0.8%, this is still relatively small, and the impact is far less than that of [O II] contamination. The small scale clustering
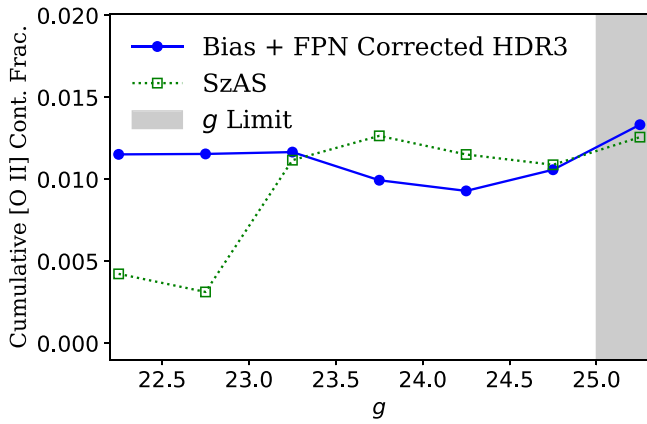
Figure 8. Cumulative (bright to faint) contamination of Lyα by [O II] as a function of g magnitude using the default ELiXer configuration. The bias + FPN corrected HDR3 curve attempts to compensate for the biases in the SzAS (compared to all of HDR3) and account for false positives in the low-S/N regime (Section 5.1).
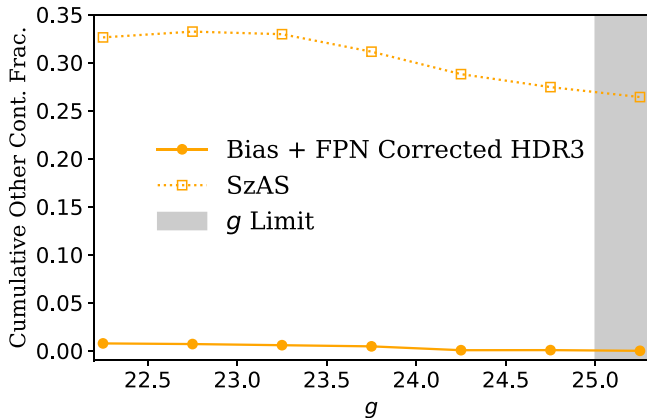

Figure 9. Cumulative (bright to faint) contamination of Lyα by emission lines other than [O II] for $g > 22$ using the default ELiXer configuration. The FPN + bias corrected HDR3 curve attempts to compensate for the biases in the SzAS and account for false positives due to random noise in the low-S/N regime (Section 5.1). The much larger contamination rate in the SzAS is largely driven by confusion of Lyα vs. C III] and C IV, where the AGN population is significantly overrepresented (see Figure 7, Sections 5.2 and 5.3).

Table 7
Cumulative Classification Performance for HDR3

Metric	SzAS	Bias + FPN Corrected
Lyα Accuracy	0.723	0.981 ± 0.034
Lyα Recovery	0.892	0.991 ± 0.033
Lyα Contamination by [O II]	**0.013**	**0.012** ± 0.001
Lyα Contamination by Other	**0.264**	**0.008** ± 0.001[a]
[O II] Accuracy	0.890	0.965 ± 0.034
[O II] Recovery	0.972	0.970 ± 0.034
[O II] Contamination by Lyα	0.039	0.021 ± 0.001
[O II] Contamination by Other	0.071	0.014 ± 0.001
Other Accuracy	0.892	0.916 ± 0.032
Other Recovery	0.509	0.294 ± 0.010
Other Contamination by Lyα	0.027	0.006 ± 0.001
Other Contamination by [O II]	0.081	0.078 ± 0.003

Notes. The cumulative performance of the ELiXer classifications on the SzAS and predictions for the full HDR3 data set for detections with $g > 22$ and using the default ELiXer configuration. The bias + FNP corrected column corrects for the sample biases in the SzAS data set and for false positives in the full HDR3 data set, assuming 30% false positive rate below emission line S/N of 5.0 and 15% rate between $5.0 < S/N < 5.5$. The values in the first column are slightly different than those in Table 6 because that table is down-selected to only consider Lyα and [O II] detections. The bold typeface rows correspond to the cumulative data points in the right-most (faintest) bins in Figures 8 and 9.
[a] 0.04 worst-case estimate. See Section 5.2 for a discussion.

of [O II] emitters projects to large-scale clustering when misinterpreted as higher-z Lyα. This is greatly diminished with C III] and C IV as the contamination sources shift to higher redshift and scales proportionally to the square of the ratio of the comoving angular diameter distances (Gebhardt et al. 2019; Farrow et al. 2021). This means the HETDEX cosmology is some $6.5\times$ less sensitive to C III] contamination than that of [O II] contamination and can tolerate ~13% (or ~16% for C IV) at the desired uncertainty. So, even the worst-case contamination is well within the required tolerances. Additionally, the focus for ELiXer has been on the largest contaminant, [O II], as the contamination rate of other lines is expected to decrease with future improvements targeting their identification.

Overall, the ELiXer accuracy is good in the HDR3 data set, while that in SzAS is poorer. The weaker performance in the SzAS set is due to the bright-magnitude and broad-emission line biases in the SzAS; this is where ELiXer does not perform as well. The stronger (estimated) accuracy in the full HETDEX population is bolstered by the large numbers of faint end detections that are highly biased toward being Lyα.

The results for ELiXer recovery rates are similarly mixed. The numbers are good for Lyα and [O II], which are, by far, the most common emission lines found by HETDEX. The recovery of all other emission lines is rather poor, and is largely an issue of the default behavior of the classification algorithms. When there is only a single line in an HETDEX spectrum, ELiXer heavily weights the various Lyα/[O II] segregation methods, which, as stated above, assume no contamination other than [O II]. In this case, ELiXer delivers a binary result, Lyα versus not-Lyα, at the expense of all other emission lines. Moreover, when analyzing particularly broad lines, ELiXer favors Lyα (often suggestive of an AGN) over [O II]; this also leads to the enhanced contamination of Lyα by such *other* lines. Additional identification metrics, such as limited run-time phot-z, spectral slope, and multi-Gaussian fits, could help improve these rates and will be explored in future versions.

A preliminary evaluation of an assessment sample expanded with ~1000 DESI-provided spectroscopic redshifts, 3/4 of which are for $g > 24$ objects, is consistent with the HETDEX classification results of this work. The resulting assessment sample more closely matches the HETDEX magnitude and emission line distributions. After the observations are complete, the full, detailed results will be presented in M. Landriau et al. (2023, in preparation).

5.3. Missing AGN and LBGs

Since ELiXer largely relies on equivalent width to classify most single-line spectra, the program currently does not perform well with Lyα emitting objects that are not classical LAEs, i.e., broad-line AGN and Lyman-break galaxies (LBGs) that may have small Lyα equivalent widths (e.g., Shapley et al. 2003). Moreover, ELiXer can also fail to find some of the broad emission lines associated with the AGN, which can result in misclassifications that would otherwise be correctly assigned

by the multiline redshift solutions (Section 3.3). This is particularly noticeable in the bright end of the SzAS (Figure 9), which has a disproportionately large number of AGN. Moreover, in AGN, ELiXer can confuse Lyα with C III] when C III] is the only significant emission line in the HETDEX spectral window ($0.96 \lesssim z \lesssim 1.25$) or with C IV when the line fit to C III] fails. Other approaches are taken to identify and recover AGN missed or misclassified by ELiXer (Liu et al. 2022), and future updates to ELiXer should improve upon its classification performance with these emission lines.

ELiXer also struggles to classify low Lyα EW LBGs. On the whole, given their name-defining detection methodology (Guhathakurta et al. 1990; Madau et al. 1996; Steidel et al. 1996), LBGs tend to be more massive and more evolved than the typical LAE (Kornei et al. 2010; Stark et al. 2010; Jose et al. 2013; Vargas et al. 2014; Steidel et al. 2018) and, consequently, may contain more dust to inhibit the escape of Lyα. While some LBGs also meet the definition of an LAE and are likely to be detected and correctly identified as such by ELiXer, the more massive objects may often be confused with low-z [O II] emitters or even overlooked completely if they exhibit weak Lyα emission or Lyα absorption. While relatively few in number compared to LAEs, the more massive LBGs do represent a highly biased mass tracer and are of value to HETDEX, so it is desirable to recover and correctly identify as many of them as possible. This means using methods that do not use equivalent width as their primary discriminant. To that end, several machine-learning approaches (both supervised and unsupervised) are being explored, as are direct enhancements to ELiXer that incorporate additional classification methods, such as run-time photo$z$ estimation.

5.4. Contamination from Noise

As stated earlier, ELiXer assumes an emission line detection is real, and not the result of noise or an artifact of the data reduction. As the S/N of an emission line detection decreases, it does become more likely that the feature is the result of noise. However, unlike real, incorrectly classified emission lines, FPNs from noise are not expected to cluster (they occur in random spectra at random wavelengths and thus map to random sky positions at random redshifts) and should only increase the uncertainty in the HETDEX cosmological measurements and not introduce a bias. As such, it is of lesser concern than misclassifications. Nevertheless, as described earlier, an (likely overly) aggressive FPNs correction (Section 5.1) is used for Figures 8 and 9 and for Table 7 to better estimate the classification performance of ELiXer against the full HETDEX data set.

Separate efforts to identify the noise-driven FPN rate include repeat observations of low S/N sources (based on the premise that random noise will not cause a repeat detection at the same position and wavelength; Mentuch Cooper 2023) and various machine-learning techniques. Their goal is to allow a more accurate model of contamination from noise.

5.5. Uncertainties

The performance of ELiXer presented in the prior sections is shown without statistical uncertainties; though some uncertainty is implicit in its predictions for the whole of HDR3.

For the SzAS results in this work, the ELiXer classifications have been taken as absolute, as the quality of the classifications has not yet been calibrated to a proper probability. (This is a planned enhancement.) Since classifications are based on votes and weights, some of which have an MCMC element with a weak dependency on the initial random seed vectors, individual executions can occasionally result in a different classification due to conditions falling just to either side of a threshold (though the quality score ($Q(z)$) is generally unaffected; see Section 3.6). Similarly the catalog-reported spec-z values are taken as truth, and matching against the reported values is done as described in Section 4, with a ± 4 Å allowance, independent of the uncertainties in the spec-z or the fitted emission line center (Section 3.1.3). Nevertheless, many realizations of ELiXer classification runs compared against the SzAS have shown the results to be highly stable and repeatable.

In projecting the SzAS results onto the full HDR3 data set, a few additional sources of uncertainty arise, such as the assumed FPN rate, which is binned only as a function of S/N. However, as with the SzAS, we assume the ELiXer classifications to be strictly categorical and the reported fractions subject only to rounding error. Anticipated expansion and improvements to the SzAS, including better matching to the HETDEX magnitude and emission line width distributions, will help address the systematics between the SzAS and the full HETDEX sample beyond the simplified corrections of Section 5.1.

As rough estimate on the uncertainties in the accuracy, recovery, and contamination rates reported for HDR3, we use the fraction of detections that are most susceptible to classification changes as described in this subsection. This is effectively captured by the largest factor in the classifications, $P(\text{Ly}\alpha)$, where $P(\text{Ly}\alpha)$ is least certain and least stable against change due to randomness in sampling (i.e., near 0.5). As 7% of HDR3 detections have $0.4 < P(\text{Ly}\alpha) < 0.6$, we assume a $\pm 3.5\%$ uncertainty on those rates.

6. Summary

As the primary emission line classifier for HETDEX, ELiXer must produce quality redshift identifications that are highly accurate, complete, and with minimal contamination. With a resolving power ranging from 750–950, HETDEX cannot split the [O II] doublet, so object classification must rely heavily on continuum information combined with equivalent width distributions. By incorporating improvements to established Lyα/[O II] separation mechanics, from the 20 Å equivalent width cut (Gronwall et al. 2007; Adams et al. 2011) to the P(LAE)/P(O II) ratio (Leung et al. 2017), and by combining additional partitioning techniques, ELiXer produces classifications that outperform the HETDEX science requirements for Lyα contamination by its principle low-z interloper, [O II] 3727 Å, while providing a good recovery rate (Table 7). The lower than required 1.2% contamination of Lyα by [O II] affords the option to loosen the project's strict classification thresholds in exchange for gains in the Lyα recovery fraction or completeness.

Although they occupy a small fraction of HETDEX emission line detections, lines other than [O II] 3727 Å, such C III] 1909 Å, and C IV 1549 Å represent a larger source of Lyα contamination in the biased SzAS. However, as described in Section 5.2, these lines are not expected to produce a significant clustering signal or bias in the $z = 2.4$ measures of $H(z)$ and $D_A(z)$. Regardless, planned enhancements to ELiXer and a larger spectroscopic redshift test sample (more

aligned with the HETDEX distribution) will improve these classifications and further reduce Lyα contamination.

The HETDEX project is continuing to work toward reducing the rate of FPN detections as a function of the emission line S/N (Mentuch Cooper 2023). Early indications suggest the contamination from noise is small above the 4.8–5.0 S/N acceptance threshold for detections. Regardless, these noise-driven FPNs should only add white noise to the LAE cluster signal. Although this increases the uncertainty in the HETDEX measurements, it should not introduce specific features in the galaxy power spectrum.

ELiXer continues to evolve. Future enhancements and revised voting criteria will be tested against expanded assessment samples drawn from forthcoming data releases. This will improve the current classification capabilities, enabling new and higher-precision science. Although ELiXer is designed for and calibrated to HETDEX, the methodology developed in this work can be adapted to other low-resolution, narrow wavelength range spectroscopic surveys.

This research benefits from the open-source projects Python (Van Rossum & Drake 2009), astropy (Astropy Collaboration et al. 2022), numpy (Harris et al. 2020), photutils (Bradley et al. 2022), and others in the open-source community.

Appendix
Example ELiXer Detection Reports

In this appendix, we include two ELiXer detection reports as examples of those used for visual inspection and diagnostics. The first, Figure 10, is a somewhat unusual HETDEX LAE: it has a very high emission line S/N, it is matched to a source contained in multiple catalogs (Section 2), it has several photometric and spectroscopic redshift determinations, and it lies in an area of sky with deep HST imaging. It is presented to illustrate the various sections within an ELiXer report. The second, Figure 11, is more representative of the typical HETDEX LAE and is shown here to that end.

1. *Summary.* From left to right, this includes (1) computed equivalent width of the emission line in the rest frame of Lyα, the combined continuum estimate (Section 3.2.4), (2) $P(\text{LAE})/P(\text{O II})$ (Section 3.4) and 68% confidence interval using the combined continuum estimate, (3) P(Lyα) score (Section 3.5), (4) quality score for the best-z redshift (Section 3.6), (5) best-z redshift, (6) classification labels (Section 3.3.4) if any, (7) error and/or warning flags[27] if any; in this example, there is a warning flag indicating a small disagreement in the g magnitudes calculated from the spectrum.

2. *Timestamp + Version.* This displays the date and time of the creation of this report and the ELiXer version number.

3. *Detection details.* Information about the HETDEX observation and the emission line detection includes, from top to bottom, (1) detection ID number and file name; (2) observation ID; (3) IFU+Amp address of the fiber nearest the detection center; (4) F as seeing FWHM in arcseconds, T as effective throughput at 4540 Å, N as dither-to-dither normalization, A as aperture correction (divisor); (5) J2000 equatorial coordinates of the PSF-weighted detection center in decimal degrees; (6) emission line wavelength center and FWHM; (7) integrated emission line flux; (8) continuum estimate (Section 3.2) from the spectrum within ±40 Å of the line center; (9) continuum estimate and g-magnitude from the full width of the spectrum; (10) equivalent width in Lyα rest frame with the continuum estimates from (8) and (9) respectively; (11) S/N and χ^2 of the emission line fit; (12) $P(\text{LAE})/P(\text{O II})$ using the continuum estimates from (8) and (9) respectively; (13) redshifts assuming Lyα and [O II]; (14) multiline emission line identification (Section 3.3), if one is selected, with its quality score, name, rest-wavelength, redshift, and equivalent width in its own rest frame using the continuum estimate in (9).

4. *2D fiber cutouts.* This is a 5×3 grid of cutouts within ±40 Å of the detection line center in the spectral direction and ± 1 fiber in the CCD direction.[28] The leftmost column is the presmoothing cutout with all rectifications and sky subtraction. The center column is the pixel flat, with any significant deviations marked in red (none in
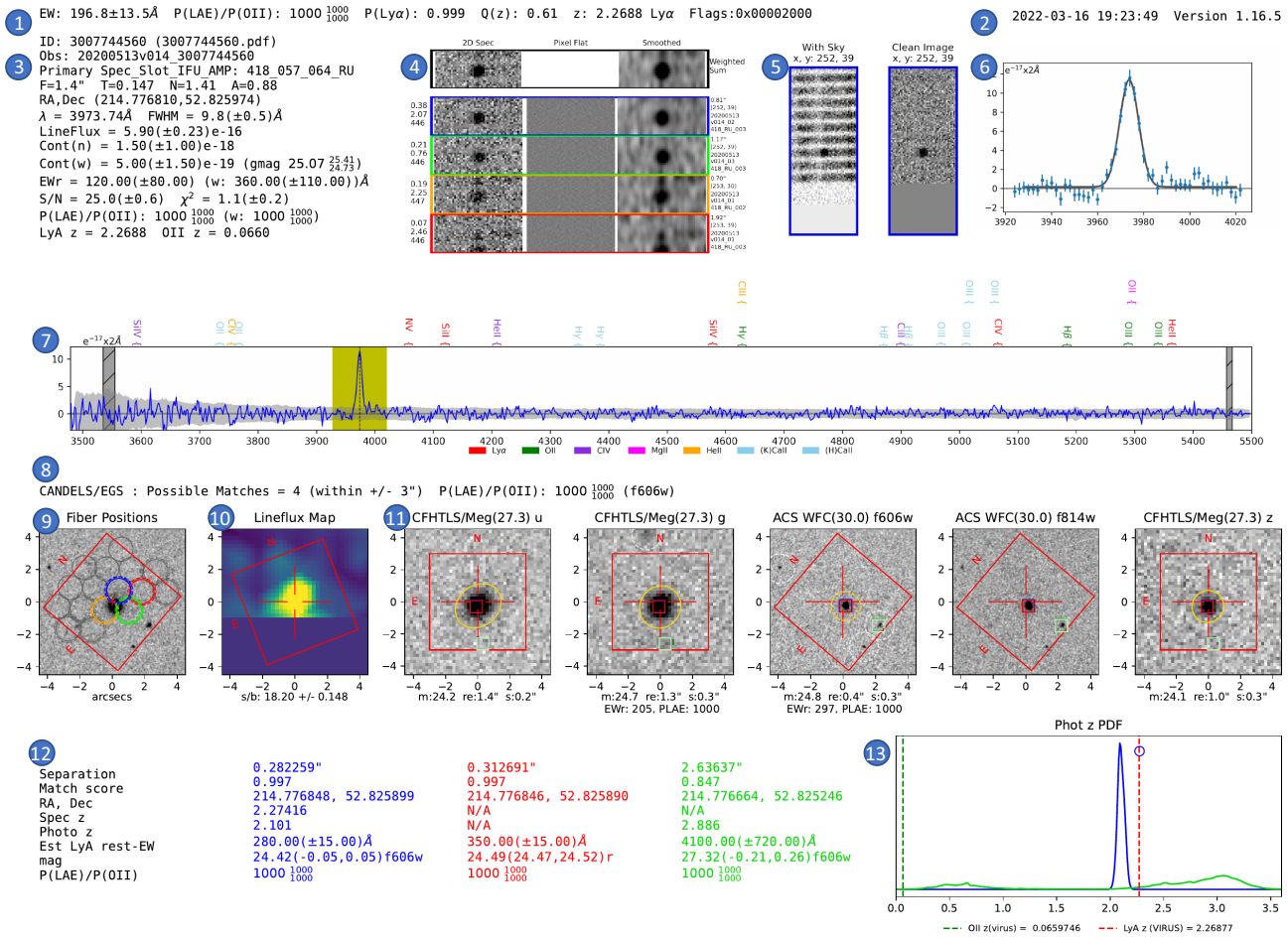

Figure 10. Example ELiXer detection report. This is a somewhat uncommon example selected to illustrate elements that are not always present for an individual detection, such as the classification label, warning flags, multiple catalog references, and photometric redshift PDFs. Descriptions of the bulleted features are provided below.

this example). The rightmost column is the same as the leftmost column but smoothed with a 2×2 Gaussian filter. The top row (highlighed in black) is the weighted sum of all contributing fibers. The rows below (blue, green, orange, red) are the highest four fibers as weighted by PSF modeled flux. The values (in very small print) to the left of the grid represent the following (from top to bottom): the normalized fiber weight in the PSF, the χ^2 of the fit to the fiber profile, and the fiber number on the CCD. The values (in very small print) to the right of the grid represent the following (from top to bottom): the fiber center distance to the detection center (in arcseconds), the CCD pixel coordinate of the fiber center, the exposure date, the observation number and exposure number for that date, and the IFU spectrograph ID, amplifier ID, and fiber number on that amplifier.

5. *Key CCD regions.* These are ± 10 fibers in the CCD direction and ± 40 Å in the spectral direction around the detection center for the fiber nearest the detection, shown before and after sky subtraction.

6. *1D line fit.* This is the 1D emission line fit to the data. This matches the gold highlighted section in the full 1D spectrum. The values are integrated fluxes in 2 Å wide bins.

7. *1D spectrum.* This is the full 1D spectrum as integrated fluxes in 2 Å wide bins. The gray background gives the

estimated. The two vertical gray-hashed bars point out the two strongest sky-lines. The gold highlighted region is the anchor emission line. Any other colored regions, if present, highlight other spectral lines that support the selected multiline redshift solution. The other red labels ("NV," "SiII," "SiIV," "CIV," "HeII") mark the positions of other possible lines in the spectrum, assuming the anchor line is Lyα; in this spectrum, none of these confirming lines are detectable. The colored labels above the spectrum represent the positions of other common lines if the anchor emission line was one of the features listed below the spectrum with the matching color.

8. *Main catalog summary.* This displays the name of the catalog with the deepest imaging used in the report, along with the number of potential catalog counterparts (if any) and the $P(\text{LAE})/P(\text{O\,II})$ found from the continuum estimate of the listed filter.

9. *Fiber positions.* This is the footprint of all fibers contributing to the detection plotted over a stacked image from the catalog with the deepest imaging. The four colored fibers match those in the 5×3 grid in (4). Fibers with a dashed outer ring are at the edge of the detector. The PSF-weighted center of the detection is marked with a red cross.

10. *Line flux map.* This is a wavelength collapsed flux intensity map summing over $\pm 3\sigma$ from the emission line
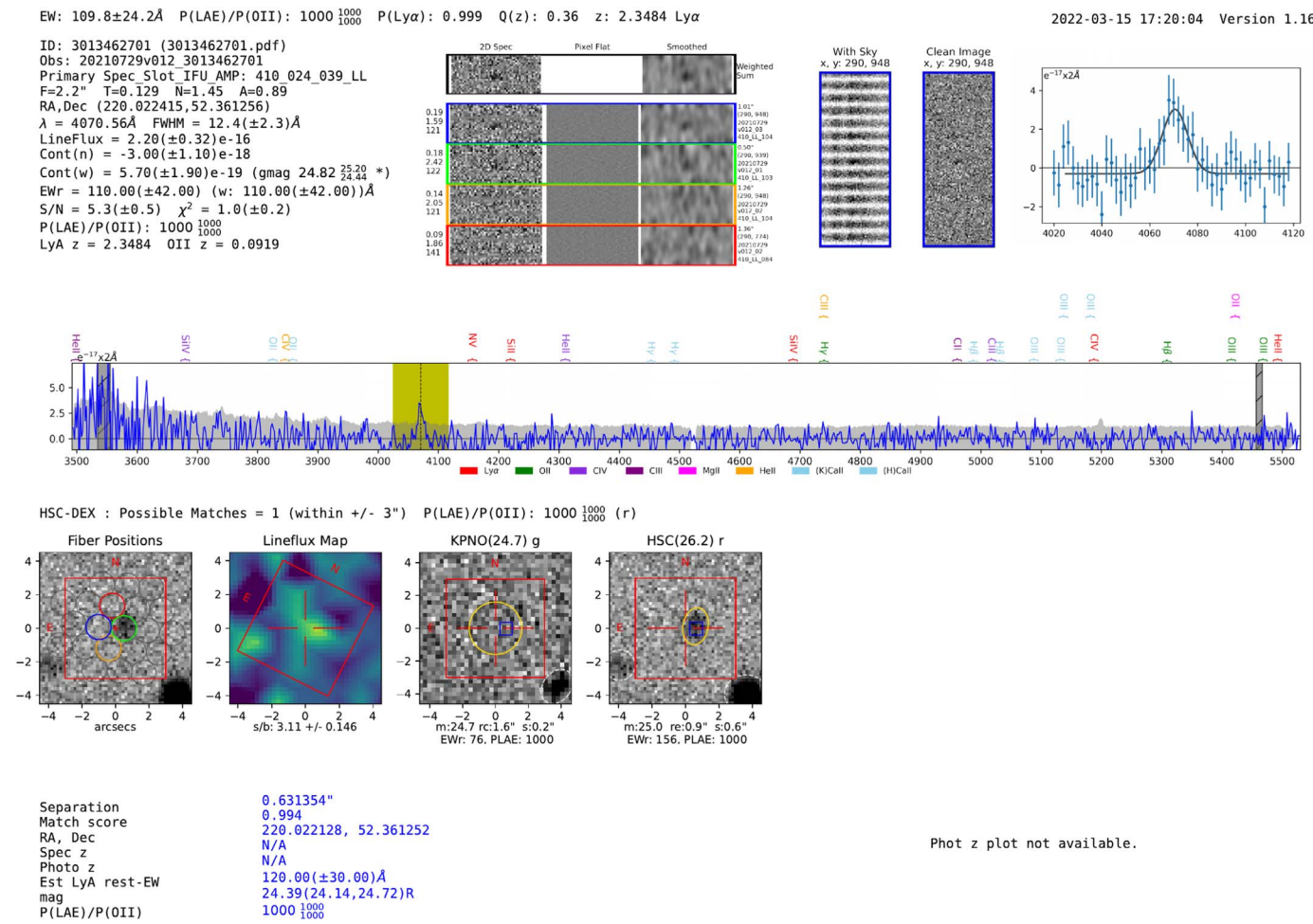
Figure 11. The ELiXer report of a typical HETDEX LAE. Note that this region of sky has fewer and shallower imaging data, and more limited catalog data compared to Figure 10. It is included here as a counter to the more illustrative, but less common example of Figure 10.

center. The values under the image are an estimate of significance based on the flux inside a $1''$ radius aperture and the standard deviation of flux inside a $5''$ to $7''$ annulus, corrected for area. The lower section of the line flux map in this example is blank as that region happens to fall off the edge of the CCD.

11. *Imaging stamps*. These are postage-stamp cutouts of the deepest imaging available to ELiXer, shown in increasing order from blue (left) to red (right). Only the bluest five filters are shown, although more may be available. The overplotted are colored $1''$ per side squares corresponding to the positions of possible catalog counterparts. The top three (see bullet (12)) are shown in blue, red, and green, with all others displayed in white. In this example, the blue and red squares overlap, so only the red is obviously visible, but they mark the same object. The overplotted ellipses are SEP identified sources (Section 2.2). A gold ellipse marks the object selected by ELiXer as the most likely counterpart, while all other objects are marked in white. If the bounding ellipse is dashed, then it has been expanded to be a $1''$ radius circle for visibility. The text above each cutout indicates the catalog name, and the approximate imaging depth and the filter. The values under the cutouts correspond to the gold aperture and are as follows: m, aperture magnitude; re, the effective radius of the ellipse in arcseconds; s, separation between the

center of the aperture and the HETDEX PSF-weighted center in arcseconds; EWr, the equivalent width in the Lyα rest frame using the aperture magnitude as the continuum estimate; PLAE, $P(\text{LAE})/P(\text{O\,II})$ using the aperture magnitude as the continuum estimate. All values are computed for *g* and *r* (or equivalent) filters, but not always for other bands.

12. *Catalog counterparts*. This includes basic information on up to the top three most likely catalog counterparts, based on magnitude and distance, which correspond to colored squares on the imaging stamps. In this example, the blue, red, and green objects are actually the same source, but reported from different catalogs, and their corresponding squares in the imaging stamps overlap. The s*eparation* is the distance in arcsec between the HETDEX detection position and the catalog-reported position. This offset can sometimes be sizeable, especially for the extended objects where the catalog reports a surface brightness center, and the HETDEX detection is more toward the object's edge. The reported $P(\text{LAE})/P(\text{O\,II})$ value uses the catalog's reported bandpass magnitude as the continuum estimate, not the aperture magnitude from the imaging stamps.

13. *Catalog z probability density functions (PDFs)*. If available, they show the photometric redshift PDFs, color coded to match the top three catalog counterparts. In this example, there is no PDF for the red counterpart

(from the CFHTLS catalog), so only blue and green PDFs are shown. Circles, again with a matching color, mark the reported spectroscopic redshift, if available. The green dashed line represents the redshift if the emission line is [O II], while the red dashed line shows the same for Lyα. Since the anchor line in this example is Lyα, there is a precise match with the spec-z and a close match with the phot-z for the object marked in blue.

ORCID iDs

Dustin Davis ● https://orcid.org/0000-0002-8925-9769
Karl Gebhardt ● https://orcid.org/0000-0002-8433-8185
Erin Mentuch Cooper ● https://orcid.org/0000-0002-2307-0146
Robin Ciardullo ● https://orcid.org/0000-0002-1328-0211
Maximilian Fabricius ● https://orcid.org/0000-0002-7025-6058
Daniel J. Farrow ● https://orcid.org/0000-0003-2575-0652
John J. Feldmeier ● https://orcid.org/0000-0003-2908-2620
Steven L. Finkelstein ● https://orcid.org/0000-0001-8519-1130
Eric Gawiser ● https://orcid.org/0000-0003-1530-8713
Caryl Gronwall ● https://orcid.org/0000-0001-6842-2371
Gary J. Hill ● https://orcid.org/0000-0001-6717-7685
Ulrich Hopp ● https://orcid.org/0000-0003-1008-225X
Lindsay R. House ● https://orcid.org/0000-0002-1496-6514
Donghui Jeong ● https://orcid.org/0000-0002-8434-979X
Wolfram Kollatschny ● https://orcid.org/0000-0002-0417-1494
Eiichiro Komatsu ● https://orcid.org/0000-0002-0136-2404
Martin Landriau ● https://orcid.org/0000-0003-1838-8528
Chenxu Liu ● https://orcid.org/0000-0001-5561-2010
Shun Saito ● https://orcid.org/0000-0002-6186-5476
Sarah Tuttle ● https://orcid.org/0000-0002-7327-565X
Isak G. B. Wold ● https://orcid.org/0000-0002-0784-1852
Gregory R. Zeimann ● https://orcid.org/0000-0003-2307-0629
Yechi Zhang ● https://orcid.org/0000-0003-3817-8739

References

Acquaviva, V., Gawiser, E., Leung, A. S., & Martin, M. R. 2014, in IAU Symp. 306, Statistical Challenges in 21st Century Cosmology (Cambridge: Cambridge Univ. Press), 365
Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, ApJS, 192, 5
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, PASJ, 74, 247
Aiola, S., Calabrese, E., Maurin, L., et al. 2020, JCAP, 2020, 047
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alam, S., Aubert, M., Avila, S., et al. 2021, PhRvD, 103, 083533
Aloni, D., Berlin, A., Joseph, M., Schmaltz, M., & Weiner, N. 2022, PhRvD, 105, 123516
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Barbary, K. 2016, JOSS, 1, 58
Barro, G., Pérez-González, P. G., Cava, A., et al. 2019, ApJS, 243, 22
Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, ApJ, 736, 31
Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
Bradley, L., Sipőcz, B., Robitaille, T., et al. 2022, Python, version 1.4.0, Zenodo, doi:10.5281/zenodo.6385735
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Ciardullo, R., Gronwall, C., Adams, J. J., et al. 2013, ApJ, 769, 83
Comparat, J., Zhu, G., Gonzalez-Perez, V., et al. 2016, MNRAS, 461, 1076
Cowie, L. L., & Hu, E. M. 1998, AJ, 115, 1319

Cuillandre, J.-C. J., Withington, K., Hudelot, P., et al. 2012, Proc. SPIE, 8448, 84480M
Davis, D., Gebhardt, K., Cooper, E. M., et al. 2021, ApJ, 920, 122
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Jr., et al. 1991, Third Reference Catalogue of Bright Galaxies (New York: Springer)
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Dhawan, S., Jha, S. W., & Leibundgut, B. 2018, A&A, 609, A72
Di Valentino, E., Mena, O., Pan, S., et al. 2021, CQGra, 38, 153001
Dickinson, M., Giavalisco, M. & The GOODS Team 2003, in The Mass of Galaxies at Low and High Redshift: Proc. ESO Astrophysics Symposia, ed. R. Bender & A. Renzini (Berlin: Springer), 324
Farrow, D. J., Sánchez, A. G., Ciardullo, R., et al. 2021, MNRAS, 507, 3187
Finkelstein, S. L. 2010, in ASP Conf. Ser. 432, New Horizons in Astronomy: Frank N. Bash Symposium 2009, ed. L. M Stanford et al. (San Francisco, CA: ASP), 33
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gallego, J., García-Dabó, C. E., Zamorano, J., Aragón-Salamanca, A., & Rego, M. 2002, ApJL, 570, L1
Gao, H., & Jing, Y. P. 2021, ApJ, 908, 43
Gawiser, E., Francke, H., Lai, K., et al. 2007, ApJ, 671, 278
Gebhardt, H. S. G., Jeong, D., Awan, H., et al. 2019, ApJ, 876, 32
Gebhardt, K., Cooper, E. M., Ciardullo, R., et al. 2021, ApJ, 923, 217
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gronwall, C., Ciardullo, R., Hickey, T., et al. 2007, ApJ, 667, 79
Gronwall, C., Ciardullo, R., Matkovic, A., Feldmeier, J., & Hay, J. 2014, AAS Meeting, 223, 246.39
Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, ApJL, 357, L9
Gurung-López, S., Gronke, M., Saito, S., Bonoli, S., & Orsi, Á. A. 2021, MNRAS, 510, 4525
Haardt, F., & Madau, P. 1996, ApJ, 461, 20
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hill, G. J., Lee, H., MacQueen, P. J., et al. 2021, AJ, 162, 298
Howell, S. B. 1989, PASP, 101, 616
Jose, C., Srianand, R., & Subramanian, K. 2013, MNRAS, 435, 368
Kirkby, D. 2020, speclite version 0.15, https://pypi.org/project/speclite/
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kornei, K. A., Shapley, A. E., Erb, D. K., et al. 2010, ApJ, 711, 693
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, Euclid Definition Study Report, https://sci.esa.int/documents/33220/36137/1567255801368-EUCLID_RB_Issue_1-1_2011-09-29HighQ.pdf, ESA/SRE(2011)12,European Space Agency
Leung, A. S., Acquaviva, V., Gawiser, E., et al. 2017, ApJ, 843, 130
Levi, M. E., Allen, L. E., Raichoor, A., et al. 2019, BAAS, 51, 57
Liu, C., Gebhardt, K., Cooper, E. M., et al. 2022, ApJS, 261, 24
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388
Martini, P., Bailey, S., Besuner, R. W., et al. 2018, Proc. SPIE, 10702, 107021F
McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, ApJ, 730, 136
Meiksin, A. 2006, MNRAS, 365, 807
Mentuch Cooper, E., Gebhardt, K., Davis, D., et al. 2023, ApJ, 943, 177
Mörtsell, E., Goobar, A., Johansson, J., & Dhawan, S. 2022, ApJ, 935, 58
Nilson, P. 1973, Uppsala General Catalogue of Galaxies (Stockholm: Almqvist)
Nilsson, K. K. 2007, PhD Thesis, Univ. Copenhagen
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Overzier, R., Lemson, G., Angulo, R. E., et al. 2012, MNRAS, 428, 778
Park, K., Matteo, T. D., Ho, S., et al. 2015, MNRAS, 454, 277
Partridge, R. B., & Peebles, P. J. E. 1967, ApJ, 147, 868
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A1
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020b, A&A, 641, A6
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Ramsey, L. W., Adams, M. T., Barnes, T. G., et al. 1998, Proc. SPIE, 3352, 34
Riess, A. G., Macri, L., Casertano, S., et al. 2009, ApJ, 699, 539
Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, ApJL, 934, L7
Saito, S., de la Torre, S., Ilbert, O., et al. 2020, MNRAS, 494, 199
Santos, S., Sobral, D., Butterworth, J., et al. 2021, MNRAS, 505, 1117
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65

Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, MNRAS, 408, 1628

Stefanon, M., Yan, H., Mobasher, B., et al. 2017, ApJS, 229, 32

Steidel, C. C., Bogosavljevic, M., Shapley, A. E., et al. 2018, ApJ, 869, 123

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJL, 462, L17

The Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270

van der Maaten, L., & Hinton, G. 2008, J. Mach. Learn. Res., 9, 2579, https://jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)

Vanzella, E., Nonino, M., Cupani, G., et al. 2018, MNRAS Lett., 476, L15

Vargas, C. J., Bish, H., Acquaviva, V., et al. 2014, ApJ, 783, 26

Verhamme, A., Garel, T., Ventou, E., et al. 2018, MNRAS Lett., 478, L60

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

Wang, W., Huang, Y., Wang, Y., & Wang, L. 2014, in 2014 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (New York, NY: IEEE), 496

Wold, I. G. B., Kawinwanichakij, L., Stevans, M. L., et al. 2019, ApJS, 240, 5